AD_____

Award Number: DAMD17-96-2-6018

TITLE: Instructional Strategies for Reducing Stress and Improving Self-Efficacy and Job Performance of Female Naval Recruits

PRINCIPAL INVESTIGATOR: Imelda S. Idar, Ph.D.

CONTRACTING ORGANIZATION: Southwest Research Institute
San Antonio, Texas 78238

REPORT DATE: October 1999

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DTIC QUALITY INSPECTED 4

20000828 216

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE October 1999	3. REPORT TYPE AND DATES COVERED Final (25 Sep 96 – 24 Sep 99)	

4. TITLE AND SUBTITLE
Instructional Strategies for Reducing Stress and Improving Self-Efficacy and Job Performance of Female Naval Recruits

5. FUNDING NUMBERS
DAMD17-96-2-6018

6. AUTHOR(S)

Imelda S. Idar, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Southwest Research Institute
San Antonio, Texas 78238

e-mail:
kgolas@swri.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

The purpose of this research project was to increase the academic success of female recruits in a technical aspect of recruit training. Two instructional interventions were developed and tested in the context of firefighting training. The Advanced Organizer intervention provided a structure for acquiring and storing technical material. This intervention orients recruits toward learning strategies that parallel their cognitive learning styles. The Role Model intervention exemplified appropriate team and individual behavior through a series of interactive scenarios that mirrored appropriate Navy behavior.

Data was collected at the Recruit Training Center in Great Lakes, Illinois with 1,235 recruits participating in the study. The results indicated that the Advanced Organizer decreased stress and made a significant positive impact on the academic success of both male and female recruits. The Role Model intervention had some positive effects on the distress and efficacy of females. The implication of these results is significant for future Navy curriculum development and for the professional development of Navy instructors. Communication and role definition should become major items in instructor professional development. Development of instructional aids such as the Advanced Organizer should be continued as an instructional approach for training technical information such as firefighting.

14. SUBJECT TERMS
Women's Health

15. NUMBER OF PAGES 174

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

X√ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

X√ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.
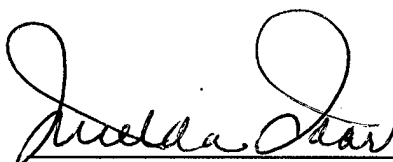
_____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

_√_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

_____   18 Sept 99
PI - Signature   Date

TABLE OF CONTENTS

i

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF APPENDIXES

APPENDIX A   SCREEN DISPLAYS FROM THE TWO INTERVENTIONS

APPENDIX B   FINAL STRESS QUESTIONNAIRE

APPENDIX C   LEARNING STYLES PRETEST

APPENDIX D   RESEARCH PAPER

TABLE OF CONTENTS

Page

LIST OF FIGURES

iv

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

A. SUBJECT OF STUDY

This study addressed methodologies and strategies aimed at improving gender integrated training in the U.S. Navy. Recognizing that the Navy operational and training environments have been male-dominated, there has been a concern that gender integration in training has not been appropriately accommodated in those efforts (i.e., learning styles, stress, self-efficacy, etc.).

B. PURPOSE OF STUDY

The purpose of this research was to identify factors which impact gender-integrated training and to test methodologies and strategies which assuage those factors. The goal of the research is to increase the academic success of female recruits in a technical aspect of recruit training where, historically, female recruits have performed less satisfactorily than male recruits.

C. SCOPE OF RESEARCH

The scope of the research included analyzing the training needs, requirements, and problems associated with firefighting training of male and female recruits; developing and testing an instrument that measured role and affective stress; and developing and evaluating two training interventions aimed at improving the academic success of recruits.

The study was designed to test three major null hypotheses:

- Providing women with models of effective performance will not reduce stress and will not improve job performance.

- Providing women with aids to structure the learning process will not reduce stress and will not improve job performance.

- Providing men and women with models of effective performance will not result in male/female attitudinal changes.

II. EXPERIMENTAL METHODS AND RESULTS

A. YEAR ONE

1. Experimental Methods (Tasks 1, 2, 3)

During Year One of this project, the research team completed tasks 1, 2, 3, and 6 of the approved Statement of Work. These tasks included analyzing the training needs, requirements, and problems associated with firefighting training of male and female recruits. The research team developed the instructional treatment plan, finalized the research design, and designed the instructional interventions. Research was conducted in the areas of cognitive science, learning theories, team competencies, communications, safety, teamwork, and leadership skills, team process measures, and gender differences in personal skills development to provide a theoretical basis for the instructional approach taken in the

1

two treatments to achieve the desired effects of reducing stress and strain and changing attitudes in male and female recruits. The following pages describe the results of the work efforts during Year One in relation to the tasks in the approved Statement of Work.

a. Analysis of Training Requirements

The team analyzed the content of the current firefighting course; observed academic and hands-on training; and collected lesson plans, slides, student guides, and other content data. The team performed a detailed analysis of the existing firefighting training at the Recruit Training Command (RTC). Instructors, recruit division commanders (RDCs), and students were interviewed to determine specific academic problem areas; desired attitudes, knowledge and skills; and male and female performance in academic tests and hands-on training. Learning objectives and programs of instruction for the academic and hands-on components of firefighting and basic seamanship were gathered and analyzed. Objectives for communication, safety, leadership, and teamwork skills were developed after analyzing RTC requirements and related research. A report describing the specific training needs in the area of firefighting at the RTC was developed and relevant data was incorporated into the Instructional Treatment Plan.

b. Development of Measures to Assess Stress, Efficacy, and Performance

Student data were developed into matrices that resulted in profiles of student mental and physical abilities. Performance assessment procedures for firefighting at RTC and at firefighting school in San Diego were reviewed and evaluated.

Additional sources of relevant information were identified and analyzed. Interviews were conducted with instructors, RDCs, and trainees to determine what factors in firefighting training gave the recruits the most difficulty. Interviews were conducted with recruits to characterize stresses, self-perceptions, and emotional reactions to firefighting training. Female and male recruits were interviewed separately and together to determine attitudes with regard to women's ability to perform in firefighting training. Interviews were semistructured, that is, a questionnaire containing open-ended questions was used to identify trainee concerns and identify RDC and instructor perceptions of stresses on male and female recruits. During the task analysis effort, the research team determined what student data were required to provide the most valid recruit profile of mental and physical abilities. The team also reviewed and evaluated performance assessment procedures for firefighting for psychometric adequacy as criteria. The initial procedures for assessing the effects of firefighting training were combined to form an Initial Firefighting Stress Questionnaire.

The steps taken to develop and revise the Initial Firefighting Stress Questionnaire were:

1. An initial 98-item questionnaire (33 stress items, 65 mood items) was constructed.
2. Confirmatory factor analyses of archival data reduced the questionnaire to 28 items. These 28 items, which measure self-efficacy and perceived stress, include 15 items from previous scales, two items formed by splitting what had been a compound item, and four new stress items based on information from recruits.
3. The Committee for the Protection of Human Subjects approved a protocol to administer the stress questionnaire to recruits.
4. A request for permission to administer the questionnaire was submitted to the Office of the Chief of Naval Operations (OCNO) as required by current policy.

2

Contingent on receipt of OCNO permission to administer the questionnaire, stress data was collected in November 1997. The stress measurement model was verified by applying measurement models. Mood measures were not included in this verification because U.S. Navy recruits provided the data used in the first year's confirmatory analyses of mood. Stress and mood measures were combined in the final Stress Evaluation Profile.

c. Development of the Treatment Plan

Upon completion of the training requirements analysis, the research team developed the Treatment Plan. The Treatment Plan provides a detailed description of how the interventions will look (the media strategy) and how they will function (the interactive design strategy). The Treatment Plan describes the research that informed the design of the treatments. The Treatment Plan describes the target population in terms of required experience level, recruit knowledge and education, and student characteristics in relation to job expectations. The Treatment Plan provides a training hierarchy list, objectives, performance factors, and qualification levels for terminal and enabling objectives. The Treatment Plan was submitted to the training specialists at RTC to ensure that the technical data is correct and that the RTC can support provision of the treatments to the recruits during the study.

d. Design of the Interventions

Once the Treatment Plan was reviewed and approved, the research team began detailed design of the interventions. The design was captured on storyboards. After completion of the draft Advanced Organizer (AO) storyboard, the research team met with the RTC instructors to validate the content. The research team also took photographs and collected videotapes for use in development of the AO treatment. The design of the AO treatment was revised to reflect the desire of the team to incorporate intelligent tutoring into the treatment. It is postulated that in addition to the learner control in the CBT, the intelligent tutoring approach will provide learning experiences that are specifically geared to the different cognitive learning styles of men and women indicated by the research. A discussion of cognitive learning styles and the relevant instructional strategies is provided in this document. The research team developed the content of the Role Model (RM) treatment with the instructors and damage control (firefighting) experts and developed a storyboard for the RM treatment.

e. Collection of Data

No empirical data was collected during Year One of the study.

f. Procedures

The procedures for Year One of the study were drawn from the systems approach to training development and the instructional system development model. Evaluation is a central function that takes place at every phase. The approach began with analysis of the training requirements and review of relevant research data which supports a particular approach to designing the instructional interventions. An analysis report was prepared which described the target audience, the learning objectives for firefighting training, and the overall training requirements. Once analysis was completed, the instructional interventions were designed and the measures for assessing stress, efficacy, and performance (the research design) were finalized. An instructional treatment plan was prepared which described the relevant research findings and the specification for the instructional interventions.

2. Results of Data Analysis (Task 1)

The results of Year One activities were the establishment of training requirements, identification and selection of relevant research findings which support the instructional interventions, finalization of the research design for the study, and design of the instructional interventions. No empirical data was collected during Year One.

a. Target Population

(1) Description of Boot Camp

Prior to fiscal year (FY) 1995, the Navy maintained three Recruit Training Commands—Orlando, Florida; San Diego, California; and Great Lakes, Illinois. Orlando was the only command to train female recruits. Until FY 94, all training was conducted in a segregated manner. The curriculum and standards were the same for both males and females. In FY 95, piloting of gender-integrated training began as efforts commenced to fully integrate the Navy. Those efforts were then concentrated at Great Lakes as it became the only site to train Navy enlisted non-prior service accessions (Navy recruits). Women are placed in either "paired" divisions, which contain one male and one female division paired for training and most other activities, or "integrated" divisions. An integrated division is one division which contains both males and females and is formed to allow recruits to participate in activities such as band and drill teams. Members of these integrated divisions tend to have higher Armed Forces Qualification Test (AFQT) scores than other divisions.

The current Recruit Training curriculum is 9.3 weeks in length. It consists of a 24-hour schedule that allows for eight hours of sleep and a stringent personal hygiene routine. The focus of the training is to transition the accession from civilian to military life. It combines a series of team building and personal skills exercises within the context of shipboard environment. The curriculum also orients and instructs the recruits on military skills such as marksmanship, firefighting, damage control, swimming, and nuclear and biological warfare. It immerses recruits in an environment which improves their physical fitness—marching, running, sit-ups, confidence course, and damage control Olympics. All exercises and instruction are conducted in a structured and disciplined manner but with positive reinforcement.

Recruit Training also functions as a screening process. Upon arrival, recruits are tested for drug and alcohol abuse. On the first day, they are put through two screens—Moment of Truth, which reviews all contract options and ferrets out inconsistencies and illegalities, and Biological Evaluation of Troops, a psychological assessment to separate individuals with severe personality disorders and/or psychiatric problems which preclude successful completion of enlistment. Recruits also undergo rigorous medical evaluation that includes testing for color blindness, hearing acuity and physical abnormalities.

(2) Characteristics of the Target Audience

At the same time that all recruit training was moved to Great Lakes and gender-integrated training was instituted, the Navy carried out an intensive review of training attrition. Attrition experienced at boot camp is especially damaging to the Navy because recruits are not reassigned but are actually separated from the Navy. The efforts were not new, having begun in FY 89. However, the Navy was directing its approach on a macro level, linking recruiting, base environment, leadership, and training factors. Also, a deeper understanding of attrition causal factors necessitated that more detailed attrition codes be used. The evolution of these codes to the current format and the myriad of training initiatives which have been executed since FY 89 often make long-term comparison of recruit demographic data difficult. For purposes of this study, statistics from the FY 97 cohort are compiled to provide a recruit profile. This

information was derived from the Navy Integrated Training Resources and Administrative System (NITRAS) and covers the period from October 1, 1996 through August 31, 1997.

During the period October 1, 1996 - August 31, 1997, a total of 43,092 accessions entered Navy RTC, Great Lakes. Of those, 86.3 percent were male and 13.7 percent were female. In general, both the male and female recruit populations are similar. The average age of the males is 20.91 and the females is 20.96. The mean education level for females is 12.11, slightly higher than that of the males, 11.94. (See Table 1.)

In order to enter the Service, accessions must take the Armed Services Vocational Aptitude Battery (ASVAB). Results are the benchmark which determine the accession's occupational direction. A subset of the ASVAB is the Armed Forces Qualification Test (AFQT). It consists of three line scores from the tests on mechanical knowledge, arithmetic reasoning, and word knowledge. The AFQT correlates positively with the psychometric construct of "g," i.e., it has a strong correlation to general intelligence. It is used as a barometer of accession quality along with the educational level. The mean AFQT for females was 59.28 and for the males, 60.63. Table 1 shows demographic data for the period from October 1, 1996 to August 31, 1997.

Table 1. Demographic Data on Accessions

	Number Enrolled	Percentage Enrolled	Average Age	Average Education	Average AFQT	Percentage Single
Male	37,197	86.3%	20.91	11.94	60.63	97.65%
Female	5,896	13.7%	20.96	12.11	59.28	95.39%
TOTAL	43,092	100%	11.97	60.45	0.13	97.34%

The female cohort had a higher minority composition than the male cohort. Of the 5,895 female accessions, 56.5 percent were white and 44 percent were minority. Of the 37,197 male accessions, 61.5 percent were white and 38.5 percent were minority. This difference has significant implications for leadership styles and communication strategies.

The Hispanic representation in both populations was the same. Among male accessions, 10.9 percent were Hispanic. Among female accessions, 10.8 percent were Hispanic. The black minority composite was significantly different. Among male accessions, 17.7 percent were black, while among female accessions, 24.7 percent were black. Table 2 shows accession data by gender/ethnic/racial groups for the period from October 1, 1996 - August 31, 1997.

Table 2. Ethnic/Racial Grouping of Male and Female Accessions

	Male Accessions	Female Accession	Percentage of Male Accessions	Percentage of Female Accessions
American Indian	1,052	206	2.8%	3.4%
Asian or Pacific	1,864	240	5.0%	4.0%
Black	6,583	1,460	17.7%	24.7%
Hispanic	4,063	639	10.9%	10.8%
White	22,878	3,303	61.5%	56.5%
Other	757	47	2.0%	0.8%
TOTAL	37,197	5,895	100%	100%

(3) Reasons for Attrition

Close scrutiny was given to attrition data, because it provides insight into the problems and strengths of boot camp. The data also gives focus to the development of training and intervention efforts. There is no academic attrition at boot camp by order of the Chief of Naval Operations. As a result, all attrition is dichotomized into two broad categories, motivational non-academic attrition and non-motivational non-academic attrition. The difference lies in the behavior exhibited by the recruit as he/she is evaluated and provided with remediation prior to separation. In motivational non-academic attrition, the recruit clearly does not want to stay in the Navy. He/she does not cooperate with staff to improve performance. In the case of non-motivational non-academic attrition, the recruit makes every effort to overcome the obstacles. Often, the recruit appeals for a waiver.

In a general review of attrition, the conclusion can be made that females are less likely to attrit. While the total recruit population experienced a separation rate of 13.9 percent, the rate was 14.2 percent for males and 11.7 percent for females. These figures are based on FYTD data provided by NITRAS for accession and attrition. From these figures it appears that females are doing better than males. However, when the reasons for attrition are examined, a telling picture emerges.

Motivational Non-Academic Attrition

NITRAS provides data on causes of attrition. For this data, attrition equals attrits divided by student flow. According to this data, 4.06 percent of male recruits attrited due to motivational non-academic causes and 5.45 percent of female students attrited for motivational non-academic reasons. The largest differences are in medical administrative attrition. For administrative attrition (failure to adapt to military life), 0.25 percent of male recruits attrited and 0.14 percent of female students attrited. Males were more overt in displaying their inability to adapt to military life. Female inability to cope was exhibited through medical problems. The data shows that 3.60 percent of male attrition was due to medical problems while 5.11 percent of female attrition was due to medical problems. (See Table 3.)

Table 3. Motivational Non-Academic Attrition

Motivational Non-Academic Attrition	Males - Percentage Of Total	Females - Percentage Of Total
1. Motivation (Negative Navy/Military Attitude)	0.01	0.00
2. Administrative (Non-Adaptation to Military Life)	0.23	0.14
3. Medical (See Table 4)	3.61	5.11
4. Legal (See Table 5)	0.19	0.16
5. Miscellaneous (Contract or Obligation - Active Duty)	0.01	0.01
TOTAL	4.06	5.45

Female attrition was almost double that of males for preservice psychiatric problems, personality disorders, and situation reaction disorders. (See Table 4.) Females tended to desert at a slightly higher rate than males. (See Table 5.)

6

Table 4. Medical (Motivational Non-Academic)

Motivational Medical	Males – Percentage Of Total	Females – Percentage Of Total
1. Orthopedic – Preservice	0.50	0.32
2. Orthopedic - Service Connected	0.02	0.00
3. Podiatry – Preservice	0.07	0.05
4. Podiatry - Service Connected	0.01	0.00
5. Psychiatric – Preservice	0.97	1.72
6. Psychological - Personality Disorders	1.21	1.51
7. Psychological - Situation Reaction	0.83	1.51
TOTAL	3.61	5.11

Table 5. Legal (Motivational Non-Academic)

Motivational Legal	Males - Percentage Of Total	Females - Percentage Of Total
1. Declared Deserter	0.03	0.04
2. Misconduct	0.16	0.12
3. PRT Failures	0.00	0.00
TOTAL	0.19	0.16

From these data, an assumption can be made that the female stress level is higher than that of males. This assumption is based on the fact that the motivational non-academic attrition was higher for females than for males and the fact that most of this attrition for females was due to psychiatric/psychological causes.

Non-Motivational Non-Academic Attrition

In general, 8.76 percentage of male attrition was due to non-motivational non-academic attrition, while 7.67 percentage of female attrition was due to non-motivational non-academic attrition. (See Table 6.) Medical and legal reasons under this aegis were similar—2.53 percent for males and 2.93 percent for females, and 0.06 percent for males and 0.07 percent for females respectively. A significant difference lies in fraudulent enlistment—5.87 percent for males and 2.22 percent for females. Males had a tendency to have higher arrest records, were more likely to be drug-dependent, and were more likely to disclose the use of drugs than females. (See Tables 7, 8 and 9.) Interestingly, females had a higher rate of drug use in boot camp than males. In the medical arena, females exhibited higher separation rates for service-connected psychiatric reasons and sleepwalking than males.

7

Table 6. Non-Motivational Non-Academic Attrition

Non-Motivational Non-Academic Attrition	Males - Percentage Of Total	Females - Percentage Of Total
1. Administrative (Hardship)	0.02	0.05
2. Medical (See Table 7)	2.53	2.93
3. Legal (See Table 8)	0.06	0.07
4. Death	0.00	0.00
5. Physical	0.00	0.00
6. Fraudulent Enlistment (See Table 9)	5.87	2.22
TOTAL NON-MOTIVATIONAL	8.76	7.67

Table 7. Non-Motivational Medical

Non-Motivational Medical	Males - Percentage Of Total	Females - Percentage Of Total
1. Dermatology (Pre-Service and Service Connected)	0.08	0.12
2. ENT (Pre-Service and Service Connected)	0.29	0.32
3. General Surgery (Pre-Service and Service Connected)	0.02	0.04
4. Gynecology (Pre-Service and Service Connected)	0.00	0.16
5. Internal Medicine (Pre-Service and Service Connected)	1.01	1.05
6. Neurology (Pre-Service and Service Connected)	0.25	0.30
7. Non-Aquatically Adapt - Hydrophobic	0.00	0.00
8. Ophthalmology (Pre-Service and Service Connected)	0.37	0.35
9. Other Medical (Pre-Service and Service Connected)	0.19	0.30
10. Psychiatric Suicide Attempts/Ideation (Pre-Service and Service Connected)	0.05	0.05
11. Psychiatric (Service Connected)	0.00	0.02
12. Psychological - Enuresis	0.13	0.04
13. Psychological - Sleepwalking	0.06	0.11
14. Urology (Pre-Service and Service Connected)	0.08	0.07
TOTAL	2.53	2.93

Table 8. Non-Motivational Legal

Non-Motivational Legal	Males - Percentage Of Total	Females - Percentage Of Total
1. Civil Conviction	0.06	0.02
2. Drug Subsequent Screen	0.00	0.05
TOTAL	0.06	0.07

8

Table 9. Non-Motivational Fraudulent Enlistment

Non-Motivational Fraudulent Enlistment	Males - Percentage Of Total	Females - Percentage Of Total
1. Arrest Record Pre-Service	0.07	0.04
2. Drug Dependent	1.18	0.50
3. Drug Disclosure	0.14	0.05
4. Homosexual Preservice	0.17	0.09
5. Initial Drug Screen (Cannabis)	3.60	1.17
6. Initial Drug Screen (Non-Cannabis)	0.69	0.37
7. Undisclosed Prior Service	0.01	0.00
TOTAL	5.87	2.22

(4) Changes to Academic and Hands-On Components of Firefighting Training

During the analysis of training requirements, the research team discovered various changes to the academic component of the firefighting course which have resulted in an overall increase in recruit scores on the academic tests, and a smaller difference between male and female scores on the academic tests. Changes to the academic component of firefighting training moved firefighting training from week four to week seven of the approximately nine-week boot camp (it is now at the end of week six). By the time they begin firefighting training, the recruits are deeper in the acculturation process and the atmosphere of basic training. Recruits indicated that by the last weeks of boot camp they had more time to study than they had earlier. They are more adapted to the demands of the physical fitness component of recruit training because they are more physically fit.

Highly technical material was removed from the curriculum and tests. Test items are regularly analyzed for suitability and changed as needed. Some recruits received academic tutoring prior to and during firefighting training from RDCs and instructors. Although changes have resulted in a greater parity between males and females in the academic portion of the course, the research team found that still more females fail the academic test. This is important because, although recruits are not attrited for academic failure, they can be set back (required to repeat the week of training) up to twice for failing the academic test. In boot camp, if a recruit fails an academic test once, he or she is allowed to repeat the test. Test result data, provided by RTC for FY 97, shows that 2 percent of male recruits and 3.4% of female recruits failed Test Four, which contains primarily items from the firefighting course, the first time they took it. (Test scores of recruits who were part of integrated divisions were not counted in this data since it is impossible to derive separate male and female scores for these divisions.) If recruits fail twice, they are "set back," i.e., they must repeat a week of boot camp. In FY 97, females were twice as likely as males to be set back for academic failure in firefighting, i.e., 0.23 percent of males failed Test Four twice and 0.59 percent of females failed it twice.

Throughout the first year of the project, the research team discovered changes to the hands-on component of the firefighting course as well. During the first visit by the research team, there was one hands-on training experience in the firefighting course, which took place in the Damage Control simulator. Recruits observed some procedures and were guided through the performance of others. There were no individual or group scores for hands-on training. On a later visit, the hands-on component of firefighting training had changed. After the guided hands-on exercise which is not scored, recruits participate in an "unguided" Team Training exercise in which they perform the same activities. In the Team Training exercise, they are supposed to act as a team without instructor intervention. There are objectives and a checklist for

9

instructors to rate each team as satisfactory or unsatisfactory on performance objectives. This is not a "test" situation. Observation showed that instructors intervened as necessary and provided immediate feedback. There are team scores, not individual scores for members of the team. There was no indication that team scores would be used to set recruits back or require them to repeat the exercise. In addition to Team Training, recruits also take part in a firefighting scenario as part of the all-night training exercise modeled on the Marine "crucible concept" training experience. Recruits do not receive group or individual scores in this exercise.

3. Research on Stress (Task 1)

Stress is a complex reaction resulting from the interaction between people and their environments. The environment can be construed in terms of a wide range of situational conditions and subjective perceptions of those conditions. People's stress includes changes in psychological status and physiological reactions. The interaction of situational demands and the person's reactions to those demands determine behavioral outcomes, including task performance in the stressful situation. Perceptions of the stressful conditions and reactions to those conditions may depend on attributes of the individual. The complexity of stress means that this concept can be operationalized in many different ways. The problem in any given situation is to identify those elements which are relevant to the specific situation.

The specific situation posed by firefighting can be isolated by combining a role theory perspective on jobs with a person-environment 6t model of stress (Kahn et al., 1964; Caplan et al., 1980; King & King, 1990). Firefighting comprises a demanding set of tasks which are important elements of the sailor's job, and instruction in firefighting techniques to meet those demands is role training. This designation of firefighting is applicable because sailors are taught to recognize the demands presented by specific conditions that occur in fires, the specific actions required to deal with those demands, and the processes of coordination with shipmates necessary for efficient teamwork. A sailor's firefighting role consists of a set of prescribed actions that should be undertaken to perform a job or specific task. Training is intended to inculcate the knowledge, skills, and experience required to perform the role.

Role stress arises from the structure of task demands. Extensive work on job stresses indicates that the stressful nature of task demands can be broadly represented by considering three potentially independent elements. Role ambiguity involves uncertainty about the specific tasks to be performed (Kahn et al., 1964; King & King, 1990). Role overload occurs when task demands exceed the person's abilities (Caplan et al., 1980). Role conflict occurs when people are capable of performing their tasks, but are faced with the need to perform multiple tasks at the same time or are expected to perform the same task different ways by different people (Kahn et al., 1964; Caplan et al, 1980). These three types of stress have been shown to apply to the analysis of basic training as an overall experience (Vickers & Ryman, 1980) and to jobs in general (Caplan et al., 1980). These concepts therefore should apply to the analysis of specific events within training or jobs, including firefighting. Two points about role stress merit special mention in the context of the proposed application to firefighting. The first point is that roles involve communication. Stresses arise in part when communications lack clarity and consistency (Kahn et al., 1964; King & King, 1990). Particularly in dynamic teamwork situations, such as those likely to occur in firefighting, the process of communication and coordination can be expected to play a key part in determining the stress level. Thus, communication problems can increase stress. The requirements for teamwork combined with situational factors of smoke, heat, and confusion that characterize firefighting may make communication simultaneously exceptionally important and unusually difficult The logical result would be exceptional role stress.

The second point is that role stresses may be correlated. For example, role ambiguity and role conflict tend to be positively correlated when people are asked to describe their jobs. The correlation is modest on

the average, but highly variable (King & King, 1990). If this trend generalizes to firefighting as a specific element of sailors' jobs, measures of the three stresses may be highly correlated. Instead of assessing several discrete stresses, the measures therefore may provide an overall index of stressfulness of firefighting. Limited discriminant validity has been noted as a possible shortcoming of current role stress assessments (King & King, 1990), but this limitation does not adversely affect the application of role stress theory to firefighting. In the present case, role stresses can be expected to show some correlation because poor communication is a common cause of different types of stress. One objective of the proposed interventions is to improve communication, thereby reducing stress. The effects of improved communication can be identified either as changes in a single general stress index, if the stresses are highly correlated, or as changes in specific stresses, if they are not. The delineation of individual stresses is an attempt to ensure reasonable coverage of the domain of stressful demands rather than an attempt to isolate individual constructs for detailed investigation.

All three aspects of role stress should be affected by the training interventions. The AO should reduce role ambiguity by defining what to do, when to do it, how to do it, and how to coordinate with teammates. This intervention should also reduce role overload either by causing changes in the person's ability level (e.g., physical training to improve strength for firefighting) or by permitting the person to develop strategies that make more efficient use of his or her abilities The latter type of effect may be particularly relevant to job performance, given that it has long been recognized that the mix of abilities required to perform jobs probably changes over time and that this change can be modeled with appropriate ability measurements (Ackerman, 1989). Finally, the AO should reduce role conflict by providing the analyses which are important to understanding how to construct more effective interventions for other applications which might be undertaken in the future.

4. Development of Instructional Interventions (Task 3)

Two instructional interventions were developed (see Appendix A for screen displays from both interventions). One of the interventions uses an RM strategy and the other uses an AO strategy. In order to design the interventions, research was conducted in the areas of cognitive science, learning theories, team competencies, communications, safety, teamwork, leadership skills, team process measures, and gender differences in personal skills development. The two interventions are computer-based and use interactive multimedia presentations. A summary of the objectives and research efforts for each of the two interventions is given below, followed by a description of the two interventions.

a. Role Model (RM) Treatment

The RM treatment was developed to help recruits develop appropriate attitudes regarding firefighting and the ability of females to perform in teams.

(1) Objectives for the RM Treatment

Our research indicates that most recruits have not actually considered that firefighting would be part of their Navy experience. Furthermore, very few know what behaviors will lead to success in situations requiring teamwork, especially in highly stressful situations. In teams, such as the fire party, they are likely to rely on strategies they know and these may not be the most effective. As the research discussed below shows, this may be especially true for females. The RM treatment incorporates human modeling techniques which were expected to influence the students to imitate the model's behavior, or more precisely, imitate the model's choice of actions. The learner will acquire an attitude which reflects the expressed or demonstrated activities of the human model.

Upon completion of the RM treatment, recruits will:

- Provide firefighting information to members of the team clearly, efficiently, and forcefully within Navy parameters.
- Use proper terminology and Navy standard communication procedures up and down the chain of command.
- Give orders with confidence, which is the product of knowledge and experience.
- State why they must learn to use protective breathing devices and firefighting equipment.
- Follow directions during firefighting exercises.
- Explain why they must refrain from inappropriate behavior, such as horseplay or unnecessary talking, during firefighting exercises.
- State why it is important to learn the safety hazards of firefighting equipment.
- Identify examples of monitoring others' performance and providing help when it is needed.
- Take and provide criticism and feedback objectively and constructively.
- Show situational awareness by identifying fire hazards.
- Show situational awareness by identifying egress routes.

To design the RM treatment, research was conducted in the areas of team competencies, communications, safety, teamwork, and leadership skills, team process measures, and gender differences in personal skills development. A summary of the research findings is presented below.

(2) Research Supporting Design of the RM Treatment

Team Competencies

Teamwork was identified as one of the competencies to be taught in the RM treatment. Since the skills that make teamwork effective are not specifically addressed in RTC, it was necessary to establish the core skills that support effective team behaviors. Research indicates that in addition to having such team skills as adaptability, flexibility, and implicit coordination (Cannon-Bowers et al., 1993), teams have a variety of other competency requirements that can be cognitive, behavioral, or attitudinal in nature (Cannon-Bowers and Salas, in press). These team competencies, which are shown in Table 10, provided a framework for identifying with the Navy subject matter experts the specific knowledge, skills, and attitudes that will develop efficacy in recruits performing in teams in the context of firefighting training.

Table 10. Team Competencies

Knowledge	• Cue strategy associations • Task-specific teammate characteristics • Shared task models, knowledge of team mission, objectives and norms • Task sequencing • Accurate task models • Accurate problem models • Team role interaction patterns • Understanding of team work skills • Knowledge of boundary spanning role • Teammate characteristics

Skills	• Adaptability: flexibility, dynamic reallocation of function, and compensatory behavior • Shared situational awareness • Mutual performance monitoring, feedback and self-correction • Leadership and team management, including conflict resolution and assertiveness • Coordination - task integration • Communication • Decision making - problem solving
Attitudes	• Team orientation: willingness to be part of a team • Collective efficacy: the team's belief that it can cope with task demands • Shared vision • Team cohesion • Mutual trust • Task-specific team work attitudes • Collective orientation • Importance of team work

Communication, Safety, Teamwork, and Leadership Skills

The skills in the areas of communication, safety, teamwork, and leadership shown in Table 11 were identified during interviews with instructors, RDCs, Navy damage control specialists, and recruits at the RTC in Great Lakes, Illinois.

Table 11. Skills Identified at RTC

Communication Skills	• Speak loudly, forcefully, and assertively • Provide brief, precise directions and information • Use correct chain-of-command address procedures • Use correct terminology • Use hand signals correctly • Establish eye contact when possible • Use the sound-powered phone correctly • Recognize the meaning of ship's alarm signals
Safety Skills	• Wear proper garb • Use breathing device(s) correctly • Handle equipment correctly • Know the chain of command and follow directions • Refrain from any inappropriate behavior • Know the safety hazards for each piece of equipment • Perform all required actions to ensure watertight integrity of the ship • Generate an escape plan
Teamwork Skills	• Know the chain of command • Follow orders • Give orders when in charge • Monitor other team members' performance • Possess team spirit • Understand the entire team function and how one's particular tasks and responsibilities interrelate with those of other team members • Provide feedback to others objectively • Take criticism objectively

Leadership Skills	• Give orders with confidence • Listen carefully to superiors and subordinates • Take responsibility for errors and correct them immediately • Make decisions quickly, based on knowledge • Pay attention to detail • Know how to talk up and down the chain of command • Establish eye contact • Repeat back to ensure accurate communication • Acknowledge and perform in accordance with orders

Team Process Variables

Glickman et al. (1987) found that two separate tracks of behavior evolve during team training. The "taskwork" track involves skills that are related to the execution of the task and/or mission. In order to train for these tasks, shared mental models are especially useful. Taskwork is supported in the development of a shared mental model of teamwork, communication, leadership, and safety within the context of firefighting training, as well as in the overview of relevant information in the AO. The RM treatment seeks to help recruits develop a shared mental model of teamwork, communication, leadership, and safety in the context of firefighting training. It is through this shared mental model that the RM treatment should affect role ambiguity and, therefore, reduce stress. (Cannon-Bowers and Salas, 1990). Teamwork involves skills that are related to functioning effectively as a team member, such as closed-loop communication, compensatory behavior, mutual performance monitoring, giving and receiving feedback, adaptability, and coordination (McIntyre et al., 1988). Teamwork knowledge, skills, and attitudes will be specifically addressed by the RM treatment. This relationship and distinction between teamwork and taskwork provides insights into the development of the treatments.

A key factor in ensuring a team's success in a cognitively complex and stressful task environment is training that incorporates explanation, demonstration, practice, dialogue, and feedback (Bailey et al., 1995). Training should focus on team performance processes which are represented by communication flow, coordination behaviors, and team strategies (Bailey et al., 1995). The four team process measures (Johnston et al., 1995), shown in Table 12, provide valuable insight into the specific skills required for effective performance in teams.

Table 12. Team Process Measures

Situation Assessment	The ability to provide communication that promotes team awareness of the surrounding environment, both external and internal to the team.
Communication	The ability to exchange information clearly and efficiently using proper terminology, standard procedures for internal and external communications, and proper phraseology.
Compensatory Behavior	The ability to monitor the activities of other team members, take action to correct team errors, give and receive performance feedback in a non-defensive manner, and provide and seek assistance or backup when needed.
Team Leadership	The ability to provide needed guidance to other team members, help team members focus their activities appropriately and anticipate tasks that should be performed, and provide instruction to other team members to enable them to perform or complete their tasks.

Gender Differences in Personal Skills Development

Tannen (1994) describes many differences in male and female communication styles as well as in ways that women and men behave as part of teams and in leadership roles. While it is useful to recognize that there are gender differences which are disadvantageous to women, it is not the goal of the RM treatment to re-adjust the Navy to female styles. The goal is to help both females and males adjust their styles to the Navy standard communication skills. It is useful, however, to consider what many women and some men will have to overcome to achieve success.

Tannen (1994) identifies a basic difference between men and women in their ways of communicating.

> Conversational rituals among women are often ways of maintaining an appearance of equality, taking into account the effect of the exchange on the other person and expending effort to downplay the speaker's authority so they can get the job done without flexing their muscles in an obvious way. Conversational rituals among men often involve using opposition such as banter, joking, teasing, and playful put-downs, and expending effort to avoid the one-down position in the interaction.

This is important to the consideration of women's role in the Navy because females' conversational rituals will work in situations where others understand the meaning of the rituals. However, in many situations, such as in the case of a firefighting team, the typical female conversation style may be ineffective. For example, in firefighting and many other military activities, there is a chain of command which, although it may be ad hoc, is expected to be followed. The nature of the task, and the need to perform in a dangerous, task-saturated environment, precludes team and leadership behaviors which may not come naturally to some women. Therefore, it is necessary for women to learn to modify their behavior in Navy team situations. Equally important, women must learn to accept other women performing in ways which are more frequently associated with males. When women use "male" conversational rituals they are perceived, especially by other women, as being domineering or abusive, while the same behavior from a male is considered strong, authoritative, and decisive. The following paragraphs describe areas where females may experience difficulty in learning to adapt to a hierarchical society in the areas of confidence, leadership, feedback and criticism, directness, and voice qualities.

Confidence. In a leadership position, recruits are expected to display confidence, which is judged by appearance and speech. According to Tannen (1994), women often appear less confident because they are more likely to downplay their certainty, while men are more likely to downplay their doubts. According to Tannen, when females make an effort to be less aggressive in their speech, they appear uncertain, even when they are not. The display of confidence was cited by the Navy subject matter experts as one of the effective skills involved in leadership. They did not say that women had more problems with the display of confidence, but this study supports the notion that it is likely that women would have more difficulty with this key aspect of leadership. The RM treatment will stress that neither men nor women should downplay their doubts, but when confidence is warranted, it should be freely expressed by both men and women.

Leadership. Leadership requires giving directions or orders to others. According to Tannen (1994), most females are acculturated against "bossiness" and have learned to suggest and give reasons. Giving short, precise orders, in a loud, forceful, confident, and assertive manner, has been identified as a necessary skill for leading firefighting teams. Tannen also provides valuable insight into the double bind in which women find themselves. If they are as aggressive or confident as men are, they are seen as bossy or worse. If they behave in a way that is more "seemly," they are often considered to be ineffective and insecure.

15

Feedback and Criticism. Tannen (1994) discovered in her research that women often give tempered criticism and expect others to give them the same type of criticism. They often experience hurt feelings when untempered criticism is given. Many men prefer straightforward criticism and they operate on the conventionalized agreement that "this is business; feelings have no part in it." Not taking criticism "personally" has been identified as a necessary skill for performing in teams. Tannen provides a valuable insight into the way women give and accept criticism. Females are often accused of not saying what they mean when they are required to provide constructive criticism. Tannen very effectively makes it clear that it is the expectations of the listener that make the message unclear. The same concept of expectations is also at the root of females' perceived inability to take criticism when it is directly proffered. The researchers were told that females need to learn to "suck it up" when they are criticized. There are many reasons posited for females' inability to accept what they consider to be harsh criticism: they don't play in rough and tumble team sports, etc. Females in military situations are unlikely to get tempered criticism from males. It is important to help them understand this need, that it is not universal, and that they can learn to accept untempered criticism in work situations while maintaining their feeling of well-being. Females are also more likely to want and expect feedback, especially praise. They need to understand that for some males in leadership roles, failure to provide feedback indicates satisfaction and confidence in them.

Directness. Tannen (1994) discovered that one area in which males and females communicate differently is directness. Those who expect orders to be given politely or indirectly may be offended. At RTC, the ability to provide brief, precise directions was identified as a required communication skill. Females were criticized for being garrulous in providing information or giving orders. Tannen's insights again can help women understand the differences and adapt to them. In adapting styles, it is necessary to consider the situation. There are some situations in which directness is absolutely necessary, such as firefighting, while in other situations, indirectness can be tolerated. In any case, it is necessary to understand that indirectness may be misunderstood by some males, and directness may be unappreciated by some females.

Voice. Tannen (1994) discovered that females often exhibit "feminine" characteristics when they speak, such as speaking softly, not projecting their voices, or speaking in a high voice. But speaking loudly, forcefully and assertively were identified at RTC as required communication skills. Tannen's research can help females to understand that there are times when a modification to their preferred style is beneficial.

Women are often in a double bind, especially in their relationship to other women. Many women prefer to be led by men rather than women. They seem to resent directness, authoritativeness, and other displays of confidence and ability from other women. The RM treatment will model behavior by women which is appropriate to their tasks, but which may be different from stereotypical female behavior.

(3) Design of RM Treatment

The content of the RM treatment was realistic firefighting scenarios in which examples of appropriate and inappropriate behaviors are demonstrated in the areas of communication, safety, teamwork, and leadership skills. Multimedia presentations with voice-over narration are used to point out specific examples of the knowledge, skills, and attitudes the recruit should develop during firefighting training.

The presentation introduces male and female sailors who are recounting their experiences fighting fires on board ships. Each sailor introduces and serves as the narrator of a short video sequence which shows examples of communication, safety, teamwork, or leadership skills. The voice-over narration identifies these skills as they are being performed.

On-line exercises are provided in which the recruit will identify compliance and non-compliance with good leadership, communication, teamwork, and safety skills.

It is expected that recruits who receive the RM treatment will develop a sense of team efficacy and a sense of self-efficacy, which will prepare them for firefighting training. It should also help them develop task-generic teamwork skills, which will stand them in good stead throughout their careers.

It was estimated that it would take the average recruit 30 minutes to complete the RM treatment.

b. Advanced Organizer (AO) Treatment

The AO treatment was developed to help the recruits learn firefighting tasks by relating the new learning to specifically relevant aspects of existing cognitive structure.

(1) Objectives for the AO Treatment

The goal of the AO was to bridge the gap between what the recruits already know and what they need to know in order to meaningfully learn to fight fires. The AO treatment provides a context of meaning for new information to be learned. It will help to orient the students toward the firefighting subject matter in such a way that the subject matter is directly related to any preexisting knowledge the students may already have. The AO will help the students anticipate the performance requirements of the job by letting them know what to expect as well as to demonstrate the desired behaviors and attitudes for acceptable job performance.

The overall objective of the AO is to provide recruits with a context of knowledge that will prepare them to succeed in the academic and hands-on components of firefighting training. Upon completion of the AO treatment, the recruits will recognize that:

- Compartment ID and closures ID are necessary for reporting and responding to a fire.
- Understanding the characteristics and uses of portable and fixed fire extinguisher systems is necessary for firefighting effectively.
- Closing ship's closures is necessary to prevent spreading.
- All sailors must know the meaning of ship's alarms.
- Knowledge of the material conditions of readiness and the use of dewatering equipment is necessary to maintain watertight integrity.
- It is important to know the roles and responsibilities of fire party personnel.
- It is important to know how to don and use all breathing devices and protective clothing.
- It is necessary to know the types of fire extinguishing materials and systems that are used for the classes of fire and to know the hazards related to each.

(2) Research Supporting Design of the AO Treatment

To design the AO treatment, research was conducted in the areas of cognitive learning styles and learner control. The research findings were used as a basis for deriving the instructional strategies on the premise that to overcome difficulties with the content of technical material, represented by the academic component of firefighting training, not only do females need analogies they can understand, they need an introduction to the information in a way that complements their unique learning styles. Computer-based training (CBT) is a dynamic medium, capable of varying such properties of a lesson as the content selected, modalities featured (sound, graphic, etc.), sequencing of topics, amount and level of practice, type of feedback, as the learners' needs dictate (Ross, 1984). The basis for program control may be as

simple as evaluation of student performance in questions. However, it is possible to develop more complex and meaningful program control based on learner characteristics. Research was performed on learner characteristics, cognitive strategies, and the implications to instructional presentation.

(3) Cognitive Learning Styles

In an effort to define cognitive style, Messick (1984) states that "cognitive learning style helps explain how an individual responds to a wide range of intellectual and perceptual stimuli." Cognitive style is not a single entity. Jonassen and Grabowski (1993) provide an understanding of the many dimensions that have been presented in the literature as comprising cognitive learning style. They identified twelve individual cognitive styles/controls: (1) reflectivity/impulsivity, (2) focal attention (scanning/focusing), (3) serialist/holistic, (4) field independence/field dependence, (5) flexibility (constricted/flexible), (6) category width (narrow/wide), (7) automization (strong/weak), (8) visual/haptic, (9) visualizer/verbalizer, (10) leveling/sharpening, (11) analytical/relational, and (12) complexity/simplicity.

Field Articulation

One area that is especially important to consider is field articulation. It is one of the few areas in which the research indicates significant male/female differences. In addition, field articulation has been studied extensively and is similar to analytic/global style, according to Ash (1986). In addressing the use of hypermedia presentation systems, Small and Grabowski (1992) discuss field articulation as a means of predicting how successfully learners will use hypermedia systems. Hypermedia is relevant here because there are many similarities between hypermedia and learner-controlled computer-based training (CBT). The literature provides insights into instructional strategies that can be employed which will make the learning experience more valuable for both field-independent and field-dependent learners. For example, field-independent individuals tend to discern figures as being discrete from their background, to focus on details, and to be more serialistic in their approach to learning. Field-dependent individuals typically see the global picture, ignore the details and approach a task more holistically. The implication for instruction is that field-independent individuals predictably would thrive on the details presented along the myriad paths available in hypermedia information systems while not necessarily gaining a global perspective of the topic at hand. Witkin et al. (1977) found that field-dependent subjects were at a disadvantage compared with field-independent subjects when required to complete a task requiring organization of material. High-spatial-ability recruits seem to have learner characteristics similar to field-independents (Shlechter, 1986).

Witkin and Goodenough (1981) characterized the information processing traits of the field-dependent style as passive. Because they have not developed sophisticated learning mediators, field-dependent individuals may use a chain-link information processing style. They also accept ideas as presented and do not modify them. They prefer teaching methods that encourage teacher-student interaction and like courses that emphasize social information. In contrast, Witkin and Goodenough saw the field-independent style as a complex and individualistic cognitive style. Its information processing traits, which include hypothesis-testing and restructuring skills, are examples of cognitive mediators that are used to reason about ambiguous and demanding problems. Their attitudes show that they prefer to learn independently and prefer courses that emphasize abstract and non-personal content.

Field dependence represents the tendency to perceive and adhere to an existing, externally imposed framework, while field independence represents the tendency to restructure perceived information into a different framework (McGee, 1979). The field-dependence/field-independence construct is also associated with personality characteristics (Olstad et al., 1981) which have instructional ramifications. Field-dependents are likely to have a more social orientation than field-independents. Field-dependents tend to seek external referents for processing and structuring their information, are better at learning

18

material with human content, are more readily influenced by the opinions of others, and are affected by the approval or disapproval of authority figures (Castenada et al., 1972). Field-independent individuals are more capable of developing their own internal referents and are more capable of restructuring their knowledge. They do not require an imposed external structure to process their experiences. Field-independents do not need external referents to process information and are better at learning impersonal, abstract material.

As characterized by Witkin (1950), field-independent learners are analytical, socially independent, inner-directed, individualistic, and possessed of a strong sense of self-identify. Field-dependent learners view things holistically and are extrinsically oriented, responsive to external reinforcement, aware of context, and cognizant of the effects that their learning has on others.

Witkin's (1981) field-dependence cognitive style theory predicts that females are more likely to have a social or field-dependent cognitive style, whereas males will more often have an analytical or field-independent cognitive style. The results of a study by Hansen (1994) indicate a significant difference between groups of students based on ethnic origin (Asian, Hispanic, White). The results of the study confirm the findings of a number of researchers who found that ethnic minority students were more field-dependent than white students (Castenada et al., 1972; Kagan and Zahn, 1975; Ramirez and Price-Williams, 1974). The implication is that instructional approaches must appeal to field-dependent individuals as well as field-independent individuals to ensure that instruction is not biased to favor individuals on the basis of gender or member of ethnic or racial groups.

Serialist/Holist

Pask (1976) identifies learning strategies as serialist vs. holist. Holists are characterized as learners who have many goals and working topics under their "aim" topic, assimilate information from many topics in order to learn the "aim" topic, entertain beliefs (often correct) about topics other than the working topics or the aim, tend to discover a global description of topics or to invent a description compatible with the conversational domain, ask questions about broad relations, form hypotheses about generalizations, and invent description schemes of their own. Serialists have one goal and working topic, which may be the aim topic; move on to another topic only when they are completely certain about the one they are currently studying; have no ideas about other topics; only describe the topic for which they are constructing an explanatory model; and ask much narrower questions.

He relates these two learning strategies to the more generally exhibited learning styles: operation learners and comprehension learners. According to Pask (1976), comprehension learners readily pick up an overall picture of the subject matter, are able to build descriptions of topics, and can describe the relation between topics. Their cognitive repertoire includes effective though individually distinctive description building operations, although such learners may not be able to apply these operations to specific subject matter information until the procedures underlying the concepts in question are specifically taught. Comprehension learners are prone, if left to their own resources, to act as holists.

Operation learners pick up rules, methods, and details, but are often unaware of how or why they fit together. They have at most a sparse mental picture of the material, and their recall of the way they originally learned is guided by arbitrary number schemes or accidental features of the presentation. If an operation learner is provided with a specific description, he assimilates procedures and builds concepts for isolated topics. The cognitive repertoire of operation learners includes accessible or effective procedure building operations and, if left to their own devices, they tend to act as serialists. Pask asserted that if the teaching strategy is matched to the same type of learning style, the recruit will learn more quickly and retain the information longer, while a mismatched condition leads to a pronounced failure to

19

understand the principles underlying the subject matter. He cautions that a disposition to adapt a particular strategy does not necessarily indicate competence in using that strategy.

Rowland and Stuessey (1988) tested their theory that a holistic approach appeared to be the best way to learn from simulations while a serialist approach seemed the best way to learn from tutorials, and found that cognitive style does interact with the mode of computer-aided instruction (CAI) to influence student achievement. They discovered that individuals who are serialists will have difficulty learning from simulations. Further, they indicate that field independence, external locus of control, high discrimination skills, and low memory skills may contribute to poor performance on simulations (Rowland & Stuessey, 1988). They recommend that for serialists, use of a simulation should be preceded by a more structured learning activity. They also recommend that holists who use tutorials should be given an opportunity to see the "big picture" before they turn their attention to the details of a subject.

Spatial/Verbal

Shlechter (1986) reports that high-spatial-ability students seem to have learner characteristics similar to field-independent students. Field-dependent people rely on external sources, such as teachers, to structure and help them understand complex perceptual and abstract stimuli, while field-independent people do not need such external aids. Field-independent people are better able to use information inherent in such situations. Instructional situations with explicit instructions, concrete presentation of stimuli, and explicit information about performance outcomes favor field-dependent students. Field-independent students are better able to perform on more abstract and less structured tasks. One would expect that computer-based training (CBT) systems with structured instructional sequencing and continuous performance feedback would be the most appropriate for field-dependent people, while CBT programs with complex graphic and learner control features would be more geared toward field-independent people.

There are many dimensions to cognitive style. The goal of this analysis was to derive instructional strategies that would apply to learners with different instructional strategies. As these dimensions were analyzed, similarities and relationships began to emerge. It became evident that males and females were more likely to prefer the cognitive style of one category over the other. Two broad categories of cognitive traits were developed and their attributes were described. (See Table 13.) In addition, Kolb's (1976) preferred learning modes (See Table 14) were looked at for their relevance to gender and instructional strategies, and Rideout's (1989) discussion of the role of gender in decision-making strategies from Myers-Briggs Type Indicator (MBTI) (See Table 15) provided a basis for the instructional strategies for the AO shown in Table 16.

Kolb's Learning Styles

Kolb's (1976) four learning styles are accommodator, diverger, converger, and assimilator. The learning characteristics of the styles are:

- Divergers excel in using imagination and brainstorming (doing and feeling).
- Convergers' dominant learning abilities are focused on finding practical uses for ideas and theories (doing and thinking).
- Assimilators are most adept at logically organizing and analyzing information, building and testing theories, and designing experiments (watching and thinking).
- Accommodators are best at learning from "hands-on" experience (doing and feeling).

Philbin et al. (1995) describe research that shows that while females are relatively well represented in each of the four learning styles, slightly more females are likely to be divergers or convergers. On the

other hand, males are much more likely to be assimilators and relatively unlikely to be divergers. Brookfield (1986) suggests that characteristics of accommodators are somewhat analogous to those of field-dependent learners, in that, like accommodators, they prefer to use trial and error methods or some variant of active experimentation in their investigation of concrete experience and they rely heavily on others for information rather than on their own analytic ability. Philbin et al. completed a study which demonstrates a significant difference in learning style, using the Kolb indicator, between the genders. They concluded that traditional educational settings may not be the best learning environment for females. Traditional education is primarily abstract and reflective, which is a comfortable fit for Assimilators. The learning style that seems to fit women the least is the assimilator. Females learn better in hands-on and practical settings, emphasizing the realm of the affective and "doing." Based on Philbin's study, females learn best when they are watching and feeling, or doing and thinking. On the other hand, males learn best when they are thinking and watching. The gender representation and preferred learning mode of each of Kolb's four learning styles are shown in Table 14 (Philbin et al., 1995).

Table 13. Attributes of Learners

Field-Dependent/Verbal Holist/Global/Comprehension	Field-Independent/Spatial/ Linear/Serialist/Operation
Field-dependent individuals typically see the global picture, ignore the details and approach a task more holistically.	Field-independent individuals tend to focus on details and to be more serialistic in their approach to learning.
Field-dependents may use a chain-link information processing style. They also accept ideas as presented and do not modify them.	Field-independents' information-processing traits include hypothesis-testing and restructuring skills.
Field-dependents prefer teaching methods that encourage teacher-student interaction and like courses that emphasize social information.	Field-independents prefer to learn independently and prefer courses that emphasize abstract and non-personal content.
Field dependence represents the tendency to perceive and adhere to an existing, externally imposed framework.	Field independence represents the tendency to restructure perceived information into a different framework.
Field-dependents are likely to have a more social orientation. They tend to seek external referents for processing and structuring their information.	Field-independent individuals are more capable of developing their own internal referents and are more capable of restructuring their knowledge. They do not require an imposed external structure to process their experiences.
Field-dependents are better at learning material with human content.	Field-independents do not need external referents to process information and are better at learning impersonal, abstract material.
Field dependence represents the tendency to perceive and adhere to an existing, externally imposed framework.	Field independence represents a tendency to restructure perceived information into a different framework.
Field-dependents typically see the global picture, ignore the details, and approach a task more holistically. Therefore, they should be less frustrated by extracting global information from the material presented.	Field-independents predictably would thrive on details presented along the myriad paths available in hypermedia presentations.
Field-dependent learners view things holistically and are cognizant of the effects that their learning has on others.	Field-independent learners are analytical, socially independent, inner-directed, individualistic, and possessed of a strong sense of self-identity.
The holist strategy uses a global approach. Learners first build broad descriptions, then fit in details. Holists learn best from simulations.	The serialist strategy uses a local approach, where the learner concentrates on narrow procedures before the overall picture emerges. Serialists learn best from tutorials.

21

Field-Dependent/Verbal Holist/Global/Comprehension	Field-Independent/Spatial/ Linear/Serialist/Operation
Holists who use tutorials should be given an opportunity to see the 'big picture' before they turn their attention to the details of a subject.	Serialists will have difficulty learning from simulations. If simulations are used for serialists they should be preceded by more structured learning activities.
Comprehension learners readily pick up an overall picture of the subject matter. They are able to build descriptions of topics and describe the relation between topics.	Operation learners pick up rules, methods, and details, but are often unaware of how or why they fit together.
Low spatial individuals are more like field-dependent learners.	Individuals with high spatial ability seem to have learner characteristics similar to field-independent individuals.
Field-dependent people rely on external sources to structure and help them understand complex perceptual and abstract stimuli.	Field-dependent people are better able to able to use information in complex and abstract stimuli.
Instructional situations with explicit instructions, concrete presentations of stimuli and explicit information about performance outcomes favor field-dependent students.	Field-independent students are better able to perform on more abstract and less structured tasks.

Table 14. Kolb's Learning Style Model

Learning Style	Percentage Of Females	Percentage Of Males	Preferred Learning Mode
Accommodator	22.0	20.0	Best at learning from "hands-on" experience (doing and feeling)
Diverger	28.9	8.0	Excel in using imagination and brainstorm, combining concrete experience and reflective observation (watching and feeling)
Converger	28.9	24.0	Dominant learning abilities are focused on finding practical uses for ideas and theories (doing and thinking)
Assimilator	20.0	48.0	Most adept at logically organizing and analyzing information and building and testing theories (watching and thinking)

Myers-Briggs Decision-Making Strategies

Rideout (1989) states that the thinking and feeling functions are the only dimensions of the Myers-Briggs Type Indicator (MBTI) in which there is a trend in gender differences. The thinking-feeling dichotomy forms the basis upon which people make decisions or judgments. Women are more likely to base decisions on feelings, whereas men are more likely to base decisions on thinking. The implication for this dichotomy is that females are motivated by appealing to their personal and subjective values, while males are motivated on the basis of abstract principles. Table 15 shows the preferred basis for decision-making by gender.

Table 15. Preferred Decision-Making Strategies

Preferred Strategy	Percentage Of Females	Percentage Of Males	Basis For Making Decisions
Thinking	35.0	60.0	Make decisions on the basis of logic or abstract principles—"the analytic factor." Contribute to society through intellectual criticism.
Feeling	65.0	40.0	Make decisions on the basis of personal and subjective values—"the bonding factor." Contribute to society through support of good works.

(4) Implications of the Research to the Design of the AO

The research conducted for the design of the AO supports the development of two broad categories of learning styles for the purposes of tailoring instructional strategies to make the instruction more meaningful to individuals. Field-dependent individuals share many cognitive similarities with verbal learners. Field-independent individuals share many cognitive similarities with spatial learners. We know that women are more likely to be field-dependent, while men are more likely to be field-independent. Research shows that field-dependent learners prefer strategies that focus on the explicit and concrete where information is presented from the specific to the general. Field-independent learners prefer strategies that focus on the abstract and unstructured where information is presented from the general to the specific.

Using the Kolb model, women are slightly more likely to be divergers and convergers, which indicates that they prefer combining concrete experience and reflective observation and finding practical uses for ideas and theories. Research also shows that men are much more likely to be assimilators, indicating that they are most adept at organizing and analyzing abstract information. Using the Myers-Briggs indicator, men have been found to prefer thinking as the basis for decision-making, while women prefer feeling as the basis of decision-making. This indicates that information given to women should stress how it contributes to the good of all, while information presented to men should stress logic or abstract principles. The paragraphs below describe the instructional strategies that will be used for each of the two broad categories.

One of the goals for the design of the AO is to present a conceptual framework for the new knowledge and skills presented in firefighting training in ways that will appeal to both "female" and "male" learning styles. During the literature review, instructional strategies that are expected to be effective for different learning styles were identified. These were applied to each of the two broad categories of learners as prescriptions for strategies that would present the material most effectively for each type of learner. This is not to say that we expect a learner to use one cognitive strategy in all situations. We strongly suspect that learners switch strategies to accommodate different subject areas and different circumstances. Therefore, the AO will provide different instructional modes to accommodate different cognitive strategies and will record which strategies are being used at particular points in the instruction, and will assess whether they are switching strategies.

The findings from the research described above were combined in order to develop the instructional strategies for the AO. The strategies are shown in Table 16.

(5) Content of the AO Treatment

The AO provides a conceptual framework for learning the following topics in the firefighting course:

1. Ship's watertight compartments and compartment ID system
2. Shipboard closures
3. Material conditions of readiness
4. Breathing devices
5. Protective clothing
6. Extinguishing agents
7. Classes of fire
8. Portable and fixed fire extinguishers
9. Firefighting party procedures and personnel

(6) Design of the AO Treatment

The AO uses a combination of learner control and intelligent tutoring with assessment of student requirements and presentation of appropriate material to provide a unique instructional experience to each student. The system will present information in ways that will specifically address the style and/or concerns of the learner.

Learner Control

The system begins by asking the student to select one of the following: "I want to learn all about firefighting training" (tutorial), "I want some information on firefighting training" (access to specific modules via a lesson map), or "I want to test my knowledge" (test). This choice is essentially between three expository approaches, a structured instructional approach (tutorial), an approach in which the learner imposes his or her own structure (lesson map), or a test. With the choice of expository approach, the learner will indicate something about his or her preferred cognitive style. It is expected that a field-dependent/verbal/linear/operation learner will select the tutorial, while a field-independent/spatial/ holistic/comprehension learner will select the lesson map. Selection of the test indicates that either the learner already knows the information or that he or she wants to determine what is already known and get an indication of what must be learned. An individual who immediately elects to take the test is likely to be field-independent.

Selection of the Tutorial. If the student selects the tutorial, this indicates his or her preference for structured instructional sequencing and a verbal overview of the topic. Since the student has indicated that he or she is a field-dependent learner, the system will structure the lesson for the student. The audio narration of the tutorial will further support the field-dependent learner by associating ideas with their practical application, providing explicit information about performance outcomes, and structuring information from the specific to the general.

Selection of the Lesson Map. If the student selects the lesson map, this indicates his or her preference to have control over the sequencing of information, thus exhibiting characteristics of a field-independent learner. Since the student has indicated that he or she is a field-independent learner, the system will provide a graphic overview of firefighting in the form of a multimedia course map. Once the introduction is over, the student will be able to select specific topics. In doing so, the student will be deriving his or her own structure of the information and exercising learner control over the structure and sequencing of the information. The course map structure fits the field-independent learning style by providing a "big picture" of the lesson material before the student's attention turns to the details of the subject.

Table 16. Instructional Strategies for the AO

Field-Dependent/Verbal/ Holistic/ Comprehension Learners	Field-Independent/ Spatial/ Linear (Serialist)/ Operation Learners
Provide structured instructional sequencing.	Provide learner control features.
Provide continuous performance feedback.	Provide necessary feedback.
Provide a verbal overview.	Provide a graphic overview with complex graphics.
Structure information to be presented.	Allow students to derive the structure of the information.
Associate ideas with their practical application.	Show (graphically) why the material is important.
Provide explicit information about performance outcomes.	Allow students to inquire about performance outcomes.
Explain how the various components fit together, then provide a global perspective.	Show the performance (global perspective) and let students identify the components.
Provide a verbal organizer.	Provide a graphic organizer.
Explain why damage control is important and how each individual contributes to the welfare of all.	Explain why it is necessary to acquire all the knowledge and skills to perform effectively.
Provide a verbal description of the "big picture."	Use a graphic presentation to show the "big picture" and let students derive the specifics.
Provide "context" for the roles the learner will perform as part of the team and what they will need to know to fulfill their responsibilities.	Provide "context" in terms of what they will learn and how they will use the knowledge to perform their roles in the firefighting team.
Use an expository presentation: Present a generality and demonstrate the necessary skills to understand the higher-level skill.	Use a discovery presentation: Allow the learner to discover the higher-level skill.
Emphasize that during training, "hands-on" experiences will be provided.	Emphasize that during training, learners will use knowledge gained in academic training.
Provide information from the specific to the general.	Provide information from the general to the specific.
Provide simulations rather than tutorials. If tutorials are used, provide an opportunity to see the "big picture" before turning their attention to the details of a subject.	Provide tutorials rather than simulations. If simulations are used, they should be preceded by more structured learning activities; or actual use of the simulation should be more structured.
During demonstrations, emphasize "feeling" aspects of activities: working toward the common good.	During demonstrations, emphasize "thinking" aspects of activities: developing expertise.

Selection of the Test. If the student selects the test, this indicates his or her preference for a discovery presentation. By participating immediately in the test, the student will discover what he or she does and does not already know. Once this is determined the student can access only those parts of the information needed. This student is exhibiting characteristics of a field-independent/holistic learner.

Learner Questions

In either the tutorial or lesson map expository approach, questions will appear on any screen that is presenting information. Prompted by these questions, the learner will be able to ask questions which will vary according to the instructional content being presented. The learner can stop the presentation at any time by asking a question.

The question will be answered and the presentation will continue. The answer is specific to the information being presented. For example, during the presentation on Compartment Identification if the learner asks, "Why is it important to know this?", the system will respond by explaining the importance of boat geography—that for safety reasons you must always know exactly where you are in case of emergencies.

In addition to answering the question, the system will also know whether the question indicates that the learner is exhibiting a particular cognitive style or that the learner has a particular concern about the material being presented. The system knows this by the type of question the student is asking. For example, if the learner asks the system, "Why is it important for me to do that?", she will be indicating a particular cognitive style, namely that she is a field-dependent learner with a concern for the "feeling" aspects of activities and how they relate to the common good. When the system determines that the learner is exhibiting a particular cognitive style, the system will present the learner with instructional content in a way that is consistent with that cognitive style.

On the other hand, if the learner asks the system, "Do women serve on damage control teams?", the learner will be indicating a concern with women's role in firefighting. When the system determines that the learner has a particular concern, it will present additional information wherever appropriate to address that concern.

Throughout the instruction the learner will retain control of the pace and presentation. The learner can switch from the tutorial, lesson map, or test at any time. It is estimated that it will take the average recruit 60 minutes to complete the AO treatment.

5. Research on Learner Control (Task 1)

Providing control of pace, sequence, content, and other elements to the learner individualizes instruction (Daniels, 1996). Accommodating the learner, as opposed to the learner having to accommodate the instruction, has been shown to result in higher achievement and improved attitudes (Frey and Simonson, 1994). Research has also shown that learner control promotes retention of information because decisions students make while progressing through instruction typically require deeper processing and reflection on the learning process. This decision-making process promotes elaboration, and allows learners to adjust the rate of encoding and processing to their individual level (Merrill, 1984; Williams, 1993). Reigeluth and Stein (1983) advocate "informed learner control." The term "informed" implies both cognitive (processes) and metacognitive (knowledge of those processes) skills.

Other findings from the research indicate that students with high ability and those with high levels of prior knowledge appear to benefit more from learner control strategies than other types of students (Williams, 1993). Subject matter also appears to influence the effectiveness of learner control treatments (Daniels, 1996). Social-oriented subjects which have less specific rules and procedures (e.g., learning communication skills) usually showed more positive results under learner control. Domains that are more rule-driven and require precise application of declarative knowledge (e.g., learning math) consistently resulted in less achievement for learner control treatments (Packard, 1996; Steinberg, 1989; Williams, 1993).

Learner control is assumed to be a beneficial feature of computer-based training (CBT) (Jonassen, 1988; Marchionini, 1988; Moore et al., in press; Park, 1991). However, it should be noted that interactive computer-based programs have the potential of creating cognitive and metacognitive problems if care is not taken during the design phase (Daniels, 1996). A commonly reported negative effect is that students can "get lost" and experience "cognitive overload" as they navigate through CBT (Chung and Reigeluth, 1992; Jonassen, 1988; Jonassen, 1991; Marchionini, 1988; Park, 1991). Sometimes when learners are

faced with navigation decisions in a complex CBT program, the cognitive demand will consume mental resources that should be available for learning (Gray, 1987; Park, 1991; Tripp and Roby, 1990). Hypermedia links that are either tangential or irrelevant cause more confusion (Nelson and Palumbo, 1992). In addition, many CBT programs do not provide navigation trails or exit paths when they are needed (Cates, 1992).

Several studies of CBT instruction have noted the absence of metacognitive ability in learners to monitor and assess their learning and take proper action to remedy their deficiencies (Jonassen, 1991; Lin, 1994; Weller et al., 1994). It is suggested that the lack of ability to assess their state of learning and take action to remedy their deficiencies leads to students missing important information, frustration, and incorrect conceptual links (Clark, 1983; Merrill, 1984; Park, 1991; Recker and Pirolli, 1992). As a result, researchers have recently begun to investigate not only the functional features of hypermedia but how these features interact with individual learners' characteristics (Daniels, 1996).

6. Research Design (Task 3)

Developments during the first year of the project affected the research design in several ways. It was necessary to change the overall design structure to bring it in line with the current organization of firefighting training at the Recruit Training Command. The frequency of stress measurements was reduced as a result of limitations on the computer resources available to the project. Proposed measures of stress were developed, then abbreviated to fit the modified research design. The assessment of performance was extended to include hands-on firefighting tests developed by the Recruit Training Command. The rationale for each of these modifications is summarized in this discussion.

a. Experimental Design

The original experimental design called for a 2 x 2 x 3 AO x RM x Hierarchical Level research design. The AO and RM elements of the design represented the experimental instructional treatments to be developed in the study. The two levels of the design for each instructional treatment were Not Exposed or Exposed to that treatment. The Hierarchical Level factor was based on the fact that people occupy different levels in the organizational hierarchy in shipboard firefighting. This initial design was to be implemented with only female recruits as the subjects of study.

The revised research design is a 2 x 2 x 2 AO x RM x Sex research design. The hierarchical level element of the original design was dropped because it is part of shipboard firefighting structure, but not part of the structure of basic training. Sex was added to the design because female recruit divisions now are paired with male recruit divisions. Paired divisions follow the same training schedule. Any attempt to separate the male and female recruits would require major changes in the standard training schedule. The effort required for this separation would substantially increase the problems the study poses for the training staff at the Recruit Training Command. Also, Recruit Training Command policy requires that any potentially beneficial instructional aids or other techniques or procedures that might improve performance be made available to both male and female recruits. Including both sexes in the study will permit an evaluation of the likely impact of the new training tools on the overall recruit population. This impact must be known to evaluate the expected payoff from implementing the new tools as part of training once the study is completed.

A further alternative to the original design has been developed to prepare for possible scheduling problems. The alternative is a 3 x 2 Treatment x Sex research design. This design has been considered to anticipate the problems that could arise if there is too little time in the training schedule to administer both treatments to a subset of the recruits in the study. Basic training is designed to cover the wide range of

topics needed to transform recruits from civilians to sailors. The objective is to accomplish this transformation with the greatest efficiency possible. The result is a very tight training schedule with little free time. In this context, the AO and RM treatments each will require a significant block of time to administer. The original research design called for some recruits to receive both treatments, while other recruits received only one treatment, and some recruits receive no treatments.

Given the tight training schedule, it may be impossible for any recruits to receive both treatments. If so, the alternative 2 x 3 Gender x Treatment design will be used. In this case, the experimental groups would consist of AO (only), RM (only), and No Treatment (control) groups. Shifting to this design would mean the loss of the opportunity to determine whether the different treatments had additive or interactive effects. This information could be important for applied decisions regarding whether to implement one treatment or both to obtain the best training outcomes. However, the revised design will provide a basis for stating whether either or both of the treatments improves training performance. The issue of combined effectiveness then may be a follow-on topic that can be investigated quickly and inexpensively if the Recruit Training Command wishes to do so.

b. Stress Measurement

(1) Stress Content Issues

A questionnaire to assess the intervention effects on stress indicators was developed and revised. The initial plan called for the questionnaire to be developed from earlier inventories, then administered to U.S. Navy recruits for assessment. A 33-item stress questionnaire was developed for this purpose. This instrument was to be combined with Profile of Mood States (POMS) to provide a complete assessment of stress effects. The POMS consists of a set of 65 mood descriptors; respondents indicate how each descriptor applies or applied to them during a stated time period.

However, a 98-item questionnaire takes a long time to administer to recruits. Past experience suggests that this would involve a period of at least 25 minutes if the questions were read to recruits. An estimated additional 10 minutes would be needed if the recruits read the questions for themselves. Familiarity with the time constraints in the training environment suggested that it would be difficult to find time to administer the initial questionnaire to recruits.

For this reason, an abbreviated version of the initial questionnaire was constructed. That questionnaire is presented in Appendix B. This abbreviation was made after review of the major study hypotheses indicated that they could be answered satisfactorily with shorter measures. The revised questionnaire consists of 28 stress questions and 12 mood questions. The 28 stress questions assess the three major stresses of interest for the project. The 28 stress questions include 15 items from previous scales. Two other items were formed by splitting what had been a compound item into two separate elements. Four new stress items were based on information from recruits interviewed about firefighting training. Seven items are designed to measure self-efficacy in relation to firefighting. As described in Appendix A, this set of items was established by exploratory and confirmatory factor analysis of data from a larger set of 26 questions. The data were obtained from U.S. Marine Corps recruits who undergo qualitatively similar experiences in basic training to those encountered by U.S. Navy recruits. The 26 questions originally were designed to measure role ambiguity, role conflict, overload, standardization, and teamwork. Standardization and teamwork were added to the item analysis because they were logical antecedents of the target stresses of interest in this project.

The final step in developing the stress assessment tools was a confirmatory factor analysis of data from U.S. Navy recruits. Permission to administer the Initial Firefighting Stress Questionnaire to U.S. Navy

28

personnel was given by the responsible Chief of Naval Operations code. The data collection was scheduled for early November and completed by the end of December, 1997.

Analysis resulted in a set of modified scales. The analysis indicated that teamwork items were of no value for measuring role conflict, role ambiguity, or overload. However, standardization items were useful. At least one item originally believed to be an indicator of standardization proved to be a viable component for each of the ambiguity, conflict, and overload scales.

The self-efficacy and perceived stress items in the Prototype Stress Assessment Profile were modified based on results of administration of the Initial Firefighting Stress Questionnaire in November, 1997. This questionnaire consists of 34 items and includes measures of role clarity, role conflict, teamwork, standardization, overload, efficacy, and additional items.

(2) Mood Assessment Issues

Mood assessment shifted from measuring the specific moods using scales from the Profile of Mood States (POMS) to measuring two general dimensions of affect. Measures of depression, anxiety, vigor, and other specific moods were replaced by scales assessing positive and negative mood.

The modified approach to mood assessment was adopted because scales designed to assess specific moods typically are moderately to highly correlated. For example, correlations between measures of depression and anxiety typically are on the order of $r = .70$. Highly correlated measures are redundant, and time constraints made redundancy a luxury that could not be afforded in this study. Relatively little information is likely to be gained by assessing specific moods in detail instead of assessing two well-established higher order mood dimensions. Furthermore, the study hypotheses do not predict treatment effects that would be evident as changes in a particular type of mood. Instead, the general hypothesis is that lowering stress will decrease negative mood and may increase positive mood or both.

The modified approach to mood assessment reduced the total number of mood items from 65 to 12. Analysis of responses to these items in previous studies of U.S. Navy recruits indicates that these brief scales should have adequate measurement precision for the study purposes.

(3) Timing of Stress Measures

The original research plan called for repeated measures of stress indicators, including mood scales. Measurements were to be obtained near the beginning of firefighting training, after exposure to the experimental treatment but before hands-on firefighting, and after hands-on firefighting. The original plan assumed that recruits would be in the computer laboratory often enough to permit these measures to be administered and recorded by computer. However, given the structured nature of recruit training, this was not the case.

Paper-and-pencil data collection was the alternative to computer data collection. Given the time required to administer a questionnaire, repeated measurements become a logistic problem for this mode of administration. Repeated interruptions in a tight training schedule could severely compromise cooperation with the project. In addition, paper-and-pencil administration involves much more personnel time not only to administer the questionnaires, but to enter the data into databases, check the data, and otherwise prepare for data analysis.

The personnel-intensive nature of the paper-and-pencil tests and the time required to administer them have shifted the testing procedure to a post-test-only assessment. This approach was somewhat less sensitive to treatment differences than would have been the case with repeated measures. Any stable

individual differences in stress remain part of the error term in the analyses, rather than being removed as a between-persons score component. However, power analyses suggest that post-test-only designs should be adequately sensitive to substantial differences between the groups. Analyses of other variables such as intelligence test scores, age, education, and other demographic attributes were added to the study to determine whether the experimental groups were composed of comparable recruits. It was reasonable to assume that the stress levels and affect were comparable prior to exposure to the treatments. This additional evaluation increased the personnel resource requirements of the recruit training component of the study.

7. Performance Assessment (Task 4)

Performance measures were obtained from standardized tests given in basic training. These scores include results of academic tests and new hands-on performance tests which have team scores. The original plan to include performance measures from post-basic training Fleet firefighting school was dropped from the study for reasons explained below.

a. Academic Tests

Test scores for Test 4 (firefighting) were obtained for FY 97. The percentage of all recruits who fail Test 4 the first time is 3.4 percent for females and 2 percent for males. The percentage of recruits who failed the test the second time, and thus were set back, is 0.59 percent for females and 0.23 percent for males. Work by other researchers has shown that firefighting academic scores are related to Armed Forces Qualifying Test (AFQT) scores (Slater, 1997).

The academic tests for firefighting had been modified. Past evidence, therefore, provided a qualitative picture of the differences between female and male recruit divisions that should generalize to the new test. The generalizable qualities of the test are that AFQT scores still should be a useful covariate, and women may continue to do slightly less well than men on the tests. The strength of the AFQT-performance association and the magnitude of sex differences in performance cannot be quantified at this time.

b. Hands-on Performance Measures

The Recruit Training Command had developed additional performance measures during 1996. The hands-on component of firefighting training is now graded on a pass-fail basis. The performance measures specify in a step-by-step fashion the actions required to fight the fires that recruits encounter in this controlled training environment. The success of the team is determined by whether or not the required actions are performed in the proper sequence. The pass-fail decision applies to firefighting teams, not to individuals.

The hands-on performance measures can be analyzed with teams as the unit of observation. Whether a team passes or fails the performance test can be recorded. The proportion of teams passing the test then can be determined for divisions that received the AO, RM, both, or neither. The proportions then can be compared using standard procedures such as the χ^2 test or tests for differences in proportions comparing the treatment groups to the control group.

Because the new procedures were in use only a short period of time, there was too little information available to determine base rates of success and to work out the number of teams that would have to be studied to permit meaningful inferences about the hands-on performance effects of the training interventions.

c. Fleet Firefighting Training

The original research plan called for a follow-up of trainees who received the experimental treatments. The objective was to determine whether the interventions affected learning and performance closer to the job site. In particular, the intent was to measure performance at Fleet firefighting training to determine whether the treatments affected the individual's readiness to master more advanced firefighting skills.

This element of the project could not be implemented for several reasons. Firefighting school adopts a different approach to learning and evaluation. In this school, the students are instructed, then asked to fight controlled fires. Evaluation takes the form of correction when mistakes are made during the procedures and debriefing to discuss how things could have gone differently. The academic tests that are given are not recorded. The firefighting itself is done as a member of a team. For these reasons, there are no firefighting scores available for individuals.

Consideration was given to developing team scoring methods. This approach was impractical because teams would include mixtures of some people who received one treatment, some who received another, and some who received both (if that option can be implemented). In addition, most or all teams would include some people who went through basic training prior to the introduction of the interventions. Thus, it would be very unlikely that any one team would consist solely of people who received the Advanced Organizer, for example. Collecting a large number of homogenous teams would be an extremely difficult task. Furthermore, the likelihood of an all-female team would be extremely small. The preceding considerations lead to the conclusion that follow-up assessment at Fleet Firefighting was not a reasonable undertaking.

d. Obtaining Recruit Training Command Measures

The Recruit Training Command significantly increased computer support for tracking and monitoring training progress in the past several years. The current system records abilities and academic test performance. These scores are readily available provided appropriate steps can be taken to obtain access to them.

The key issue was whether individual identifiers would be needed to pair the data from the academic files with those from other sources. Analyses of the general effects of interventions could proceed without the individual identifiers. Recruit divisions are likely to be the unit of sampling for treatment administration. Thus, the data could be extracted for all individuals within a division without asking for individual identifiers. If roughly the same proportion of recruits from each Division attend the intervention sessions, this approach would provide treatment estimates that were roughly comparable for comparisons between interventions. The comparisons would be biased downward slightly when comparing treatment and control groups because some in the treatment group would effectively be controls (i.e., will have missed the session and, therefore, not be exposed to the treatment). The exact magnitude of this problem was studied in more detail as the treatments were implemented.

e. Psychometric Evaluation of the Recruit Training Measures

A technical psychometric evaluation of the academic performance data was not appropriate. Such evaluations typically focus on considerations of measurement precision (i.e., reliability), factor structure, and other similar criteria. These criteria are relevant to the measurement of psychological constructs which involve a single dimension of individual differences (e.g., depression, stress). In these cases, it is reasonable to assume that response differences between individuals reflect differences in the strength or quality of the underlying psychological processes that give rise to the behaviors and feelings referred to in the items. The items are chosen to reflect those underlying processes, so some degree of coherence (i.e., correlation) among the indicator variables is expected. The absence of that coherence is reason to question the validity of the scale.

Academic achievement measures do not fit the standard psychometric model, except in special cases. Suppose, for example, that a course involved learning several different principles and their applications. The probability of getting the correct answer to two different questions then will depend on whether the questions draw on the same principle. If different people master different principles, there may be little correlation between questions involving different principles. On the whole, the correlation between getting one question right and getting another question right may be quite small because different principles were involved in many item pairs. For this reason, standard psychometric models are of limited value in evaluating academic tests. In short, the validity of an academic test does not rest on internal consistency of the measure.

Content validity is the major concern for an academic test. Does the test accurately reflect mastery of the material that is being taught? If so, the test is valid in the most important sense given the purpose for which it was constructed. The content validity of the academic tests has been reviewed extensively by U.S. Navy educational experts within the past two years to ensure that it maps onto current course content. The nature and content of questions and the instructional content of the firefighting course have been reviewed to ensure that the training meets the objective of preparing recruits to fight fires in the fleet. This review was conducted by U.S. Navy subject matter experts who have endorsed the existing test as an accurate representation of the academic content of the firefighting course. Based on this review, the academic tests are valid indicators of mastery of the knowledge required for firefighting.

8. Conclusions and Recommendations from Year One (Task 6)

Upon conclusion of Year One of the study, a recommendation was made to continue to implement the Statement of Work (SOW) as proposed. There were no anticipated changes to the SOW. However, the following changes were made to the research design. An explanation of each change and the reasons for the changes is provided in the discussion above.

- Replacement of gender for hierarchy in the research design
- Change in the number of stress measures from three to one post-test measure
- Deletion of the plan to include performance measures from Fleet firefighting school
- Change from computerized to paper-based stress measure

After the first year of the project, no definitive conclusions were derived from the application of the yet-to-be-completed development of interventions. However, all of the objectives for the first year were met: analysis of the training requirements; development of measures to assess stress, efficacy, and performance; development of the Treatment Plan; and design of the multimedia interventions.

32

The recruit demographics, attrition analysis and literature review strongly support the assumptions that were made at the inception of the project. Females are a minority in the Navy population. In Recruit Training, they comprise 13.6 percent of that population. In the Navy force structure, they account for 15 percent. Navy historically has been a Service with strong male-dominated traditions. Its current efforts to fully integrate the females must also include interventions which allow both female and male sailors to recognize and accommodate leadership, personal, and communication strategy differences. This research should provide greater insight into those differences. Better understanding allows for an improved ability to address these differences with sound, theoretically valid pedagogical interventions.

The literature review, particularly the studies by Tannen (1994, 1990) and Gray (1992) supported the Navy experience in training. These authors gave names to issues that had been intuitive to Navy instructors—females seemed to be more emotional, leadership styles were different, communication strategies did not match a male-dominated Navy lexicon. Recognition of those differences is the first step in improvement of instructional methodologies and content. A corollary that warrants further investigation and will be addressed in the project is, "Are there gender differences in cognitive learning styles?" Attrition data seem to support that concept. Females are experiencing a significant amount of stress in the recruit training environment. For reasons such as personality disorders, psychiatric factors, and adjustment disorders, females attrite at almost double the rate of males. There are several factors which account for this phenomenon. Leaders treat females differently and are more likely to provide psychological help to women than men. Women tend to be more open and to discuss their emotions more than men. Finally (the focus of this project), women are experiencing more stress in the male-dominated training environment. The project includes the identification and development of stress assessment tools to measure the impact of the project interventions on the female recruits.

B. YEAR TWO

1. Experimental Methods (Tasks 4 and 5)

During Year Two of the project, the research team performed tasks 4, 5, and 6 of the approved Statement of Work. The procedures for Year Two of the study were drawn from the systems approach to training development and the instructional system development model. Described below are the statistical analyses performed to analyze the data, the procedures for development and evaluation of the interventions, and the procedures for revision of the Stress Profile questionnaire.

a. Statistical Analysis Methods

The following statistical analysis methods were performed.

1. Independent sample t-tests were performed to evaluate female-male differences on pre-training variables.

2. A one-way analysis of variance (ANOVA) with treatment (AO, RM, control) as the group classification variable was performed to assess the effectiveness of the randomization procedures. The dependent variables were Armed Forces Qualification Test (AFQT) and pre-training stress measures.

3. Paired sample t-tests were performed to describe changes in stress levels during firefighting training.

4. Analyses of covariance were conducted to provide an overall test of the effects of gender and treatment on stress and performance. Gender (female-male) and treatment (AO, RM, control) were the grouping variables in these analyses. These variables were treated as fixed factors. The covariates in the analysis depended on the dependent variable being analyzed: AFQT for academic performance and pre-training scores for role ambiguity, role conflict, self-efficacy, positive mood, and negative mood. A separate analysis was conducted for each criterion variable.

5. Residualized scores were computed for each dependent variable to analyze patterns of change in more detail. The residualized scores were computed by regressing academic test scores on AFQT and each stress indicator on the corresponding pre-training score. The residualized score was the difference between the actual post-training score and the predicted post-training score. Treatment effects then were computed by determining the average residualized scores for specific groups (e.g., females receiving the AO treatment). The analyses of residualized scores included:

   - *Analyses of variance to provide focused tests of the effects of the AO and RM treatments.* In each case, a 2 x 2 ANOVA was performed with gender as one group classification variable. The other group classification variable was an AO contrast or an RM contrast. The AO contrast was constructed by combining the RM and control groups into a single group with the AO group as the second group. The RM contrast was defined the same way, except with RM in place of AO. Where appropriate, these ANOVAs were followed by t-tests for independent groups to clearly isolate which groups differed from which others.

   - *Pearson product-moment correlations between the residualized performance score and the residualized stress scores.* The residualized performance score was an index of whether the person performed better than expected based on his or her ability (positive residual) or worse than expected (negative residual). The residual stress scores indicated whether post-training stress was higher than expected based on anticipatory scores (positive stress residual) or lower than expected (negative stress residual). Exceptionally high stress would be associated with poorer performance if the correlations were negative for role ambiguity, role conflict, and negative mood. In these cases, a positive residual implies high stress, so a negative correlation links high stress with lower than expected performance. In the cases of self-efficacy and positive mood, negative correlations would indicate that poor performance was associated with stress. This reversal occurs because lower self-efficacy and positive mood indicate stress.

b. Development Methods for Interventions

During Year Two of the project, the interventions were developed and the Stress Profile questionnaire was revised. The methods for development of the treatments consisted of the following tasks.

1. The AO and RM storyboards (design documents) were reviewed and approved by the Navy.

2. The lists of required media elements (graphics, video, photographs, audio) to effectively present the content were generated for the two interventions.

3. Existing videotapes were obtained from the Navy and added to those obtained at the Recruit Training Center (RTC) at Great Lakes Naval Base to incorporate realistic footage where possible in the AO and RM interventions. The incorporation of existing video added to the realism of instructional scenes. Existing video footage (approximately 14 videotapes) was reviewed for content. Specific frames were identified for use primarily in the AO.

34

4. Basic database structures and variables were designed to gather user interaction data for both instructional interventions. Authorware was chosen as the programming language for the computer-based interventions and the data was saved as delimited text files that could be viewed in several types of software. (After the evaluation, Microsoft Access was used to organize the data for statistical analysis.)

5. During a two-week time period, Southwest Research Institute (SwRI) and the U.S. Navy video production crew videotaped content for the RM treatment at Great Lakes Naval Base. Two subject matter experts (SME) from the RTC were present for scene development and review. A damage control (DC) instructor (male), a DC-rated sailor (female), and approximately 10 recruits served as the talent. In addition to the former two sailors, six other sailors (petty officers) were also videotaped for RM testimonial content. SMEs performed final review of AO and RM instructional storyboards for content and accuracy.

6. The Navy video crews edited video segments and sent footage to SwRI for further identification and digitization.

7. Video footage and still photographs were digitized for use in the treatments.

8. Professional audio narration was recorded and digitized for both treatments. Three audio recording professionals, two females and one male, were chosen to record the narration for the AO treatment, and one female narrator recorded the audio for the RM treatment.

9. Instructional graphics for inclusion in both interventions were produced.

10. The graphical user interface (GUI) was developed by the graphic artist and treatment development teams.

11. Treatments were programmed in Authorware.

12. Database variables were incorporated into the programs and tested for accuracy in gathering data.

13. SwRI revised the consent form and received approval from the Institutional Review Board (IRB). Revisions to the consent forms included making the form easier to read and designing it to fit on one page.

14. SwRI developed the procedures and checklists to support data collection for use during the evaluation.

15. The SwRI team traveled to the RTC at Great Lakes Naval Base for implementation of treatments.

16. The Navy computer lab was prepared for data collection. Computers were tested and "fixed" (sound cards, headphones, video replay, etc.) by Navy personnel. Treatments were installed on 46 available computers in the Fundamental Applied Skills Training (FAST) Learning Resource Center (LRC). AO was installed on 23 computers and RM was installed on 23 computers. Treatments were identified on computers labeled "AO" or "RM".

17. Implementation procedure checklists were followed during implementation to ensure consistency during the data collection effort.

18. The treatments were revised in response to data generated during the pilot evaluation.

c. Revision of Stress Profile Questionnaire

Revision of the Stress Profile questionnaire was accomplished during Year Two of the study, prior to implementation of the interventions. A decision was made to repeat some elements of the prior exploratory analyses for the measurement structure of stress. This decision was made because some items used to measure the stresses of general recruit training did not adapt well to the specifics of firefighting training. In addition, some items were modified significantly (e.g., one ambiguity item was split into two distinct items) and three new items were added. These changes were sufficient to make it reasonable to replace the planned confirmatory analyses with more exploratory analyses.

d. Procedures for Implementation of Interventions and Data Collection

The procedures for implementation of the interventions and data collection consisted of the following tasks.

1. The data collection effort began with administration of the consent forms and Stress Profile questionnaires to the recruit participants. The recruits completed the paperwork within one hour, in a classroom located in the RTC Headquarters building (178) by division. Two integrated recruit divisions, 233 and 234, were in the classroom at the same time. Division 233 consisted of males and Division 234 consisted of females.

   Recruit participants were introduced to the administrators and topic of the study. The consent forms were distributed and read to the recruits. Questions were answered, and the administrators instructed any recruits who did not wish to participate in the voluntary study, or who had extenuating circumstances, to move to another area of the classroom. (Extenuating circumstances included evening watch duty conflicts and other schedule conflicts.) The consent forms were signed by the participating recruits and collected. Of a total of 349 recruits in the divisions, 271 participated in the study. Seventy-eight (78) recruits did not participate due to extenuating circumstances.

2. After completion of the consent forms, the Stress Profile questionnaires were distributed to all the recruits who would participate in the study. Directions were provided on how to fill out the scantron forms to answer the Stress Profile questions. Recruits were shown how to fill in their Social Security number, gender, division, name, and questionnaire number (1, 2, or 3). The first student in the first row was given number one, the second student was given number two, the third student was given number three, the fourth student was given number one, and so forth. All of the students who were given the number one received the AO treatment. The students assigned number two received the RM treatment. The students assigned number three became the control group. The Stress Profile questionnaire was explained and students were instructed to answer the questions with regard to how they felt about recruit firefighting at that moment in time.

3. In order to randomly assign recruits to treatment or control groups, after the recruits completed the Stress Profile questionnaire, the scantron forms were divided into three groups in relation to the questionnaire number indicated (1, 2, or 3). The scantron forms were counted to ensure equal cell (sample) size among treatment and control groups. Group 1 was assigned to the AO group and Group 2 was assigned to the RM group. In order to accommodate the number of computers available in the lab and the number of students receiving treatments, these two groups were instructed to return that evening at a designated time (5:40 pm or 7:40 pm) depending on their division. For example, Groups 1 and 2 from one division returned at 5:40 pm while Groups 1 and

2 from the other division returned at 7:40 pm. Recruits in Group 3 were instructed that they would not be taking the computer-based exercises. They were ensured that they would receive all of the same training that all other recruits received at the RTC. Group 3 recruits are the control groups for the data analysis.

4. As the recruits arrived at the FAST lab, they were instructed to sit at computers designated by a sign that read "AO" or "RM" depending on the treatment group they were assigned to. For example, recruits in Group 1 were instructed to sit at the computers labeled "AO" and recruits in Group 2 were instructed to sit at the "RM" computers. The administrators told the recruits that there were two different computer-based applications in the lab and that they would interact with only one of them. The administrator explained that there might be a difference in the amount of time it would take them to complete the programs, based on the application and on the individuals. The administrators explained that the recruits should take their time while interacting with the program, as they had ample time (90 minutes) within which to finish. As recruits completed the program, they were directed to a classroom to wait until the entire division was finished.

    Recruits were encouraged to ask questions and make comments to the administrators at any time during the program. The recruits were told that we were evaluating the system and any feedback from the recruits would be helpful toward improving the system for future uses.

5. The recruits completed the treatments. The computer screens started at the Windows 95 desktop interface with the "sign on" application highlighted. The recruits were shown how to click with a mouse and enter data using the keyboard. The administrators instructed the students as a group through the sign-on application. After the recruits were "signed on", they were instructed to begin the treatments. The administrators emphasized how important it was that the students not quit the program, because all data would be lost and would not be recoverable.

6. During the students' interaction with the program, they were reminded how much time they had left beginning at the 60-minute point. The time left was announced at 30, 20, 10, and 5 minutes. When all recruits had completed the programs and were seated in the adjacent classroom, a cross-check was performed on the names in each group and the actual treatments the students had completed. Each division had one or more persons who had sat at the "wrong" computer and were given the "wrong" treatment. This data was collected, which explains the differences in cell sizes between treatments and control groups.

7. After dismissal of the division in its entirety, the administrators prepared the lab for the next division's arrival. Administrators returned all computer screens to the Windows 95 desktop interface and inserted fresh program diskettes into the floppy drives for collection of the data for the next group of students.

8. The same procedures were followed for the next group of recruits. After dismissal of the division, the paperwork was filed, floppy disks were collected, and the computers were reset to the Windows 95 desktop interface for daytime use in the lab.

2. Results of Data Analysis (Task 5)

a. Stress Measurements

Exploratory principal component analyses were conducted for the 40 stress items. These analyses were adopted in preference to the confirmatory factor analyses planned at the time the project was initiated. The exploratory analyses were adopted as a first step because a number of items had been included in the study that were not originally designed to measure either role ambiguity or role conflict. Some new items were added to reflect comments from recruits about firefighting training. The additional items could affect the overall structure of measurement in ways that would not be evident if a selected set of models was imposed on the data. On the other hand, if the original measures of role ambiguity and role conflict retained their structure even with these changes, the exploratory approach would still identify them. The exploratory approach, therefore, provided an opportunity to identify points at which the intended measurement model broke down, if it did, but would still recover that measurement model if it was appropriate.

The initial model provided a reasonable basis for measurement. The structure of the item set indicated the presence of more than two stress dimensions, but the close approximations to the original scales were recovered. First, the analyses indicated the presence of five eigenvalues >1.00 for both the pre- and post-training data. However, in each case, only three values were greater than expected by chance according to the Monte Carlo results of Cota et al., (1993). Therefore, three components were extracted.

Two of the three extracted components were close approximations to the *a priori* role ambiguity and role conflict scales. This conclusion was reached by rotating the extracted components to an orthogonal varimax solution. The rotated components were matched across solutions by determining which three items had the highest loading on each component. After matching components, an item was designated as an indicator for a component if: (a) the component loading was > .30 in both the pre-training and post-training data, and (b) loading on other components was < .30.

Role ambiguity and role conflict scales were constructed by combining the original item classification with the above criteria. Items 5, 12, 13, 16, 23, 24, 25, and 28 met these criteria for role ambiguity. Seven of these eight items were designed to measure role ambiguity. Item 5 initially was designed to measure overload (Vickers & Ryman, 1980) and was included in this study as a potential indicator of role conflict. That interpretation of the item responses assumed that conflict was generated by having to choose between one of several tasks that needed to be done at the same time. In the context of firefighting, however, this item appears to measure the ability to get work done quickly because roles are clearly specified.

Items 2, 9, 15, 20, and 22 met the scale construction criteria for role conflict. These items comprised five of six items included in the analyses that originally were designed to measure role conflict.

The third component in the analysis was defined by items originally designed to measure several distinct constructs. Items that met the scaling criteria for this component originally were designed to measure role ambiguity, role conflict, teamwork, and standardization (Vickers & Ryman, 1980). The set of items lacked any clear unifying theme, and component loadings tended to be inconsistent from the pre-treatment to post-treatment period.

No scale was constructed from the items defining the third principal component in the analyses. The items with substantial loadings were primarily those included in this study in an attempt to expand the boundaries of the role ambiguity and role conflict concepts in the context of firefighting training if

appropriate. Instead, the items appear to have been useful in producing narrower, more highly focused scales for role ambiguity and role conflict. A scale constructed of the items defining this component would have had no meaning in the context of the specific theoretical focus of the study. The inconsistent component loadings raise concerns about the appropriateness of constructing a scale from those items. Even if a scale had been constructed, it would have had no obvious relationship to the study hypotheses. Those hypotheses dealt with role conflict and role ambiguity. Acceptable scales for conflict and ambiguity had been identified, so the focal stresses for the research were covered.

The principal component analysis (Table 17) produced scales suitable for assessing role ambiguity and role conflict as these constructs occur in the context of firefighting training. The role ambiguity construct is measured by items that indicate that responsibilities are clearly spelled out. Descriptive statistics for the pre- and post-training scales are given in Table 18.

Table 17. Principal Components of Stress Items

Item No.	Component:	1		2		3	
	Time:	Pre	Post	Pre	Post	Pre	Post
	Role Ambiguity						
13	Procedures detailed	.75	.68	.18	.30	−.03	.14
28	Orders clear	.69	.67	.07	.10	.04	−.15
12	Communications clear	.67	.67	.32	.27	.07	−.05
24	Only one way to do each task	.48	.63	−.02	−.04	−.07	.08
23	Team responsibilities clear	.75	.62	.07	.21	.04	−.26
25	Goals clear	.62	.57	.21	.31	-.07	−.07
16	Explanations clear	.57	.61	.38	.36	.11	−.04
5	Recruits work fast	.39	.34	−.17	.33	.28	.02
	Role Conflict						
15	Rules often conflict	.07	−.15	.10	−.01	.65	.73
9	Tasks done by numbers	−.11	−.21	.22	.05	.74	.69
20	Confusion over who does which tasks	.29	.01	−.22	−.18	.54	.64
22	Difficult to keep up	.17	.07	−.37	−.26	.46	.61
2	Done differently	.07	.00	.27	.13	.53	.34
	Third Component						
7	Know expected standards	.21	.41	.72	.42	.11	.00
3	Recruits cooperate	.51	.19	.27	.71	.04	-.03
4	Recruits follow standard procedures	.40	−.05	.40	.70	−.14	−.02
1	Know which procedure	.11	.21	.75	.59	.18	−.07
10	Remember technical details	.33	.34	.52	.39	.05	−.06

Item No.	Component: Time:	1 Pre	1 Post	2 Pre	2 Post	3 Pre	3 Post
	Not Included						
8	Conflicting orders common	.32	.26	.24	.48	.28	.00
17	Leader's decisions good	.69	.23	.02	.58	.12	−.09

Note: See text for description of principal component analysis and component matching procedures. "Pre" and "Post" column headings refer to pre-training and post-training data.

Table 18. Descriptive Statistics for Stress, Efficacy, and Distress Scales

Scale	Pre-Training			Post-Training			r_{tt}	t-test	Sig.
	Mean	SD	Alpha	Mean	SD	Alpha			
Role Ambiguity	2.94	0.64	.80	2.20	0.57	.79	0.36	18.01	.000
Role Conflict	3.00	0.61	.56	2.68	0.71	.57	0.21	6.43	.000
Self-Efficacy	3.78	0.56	.58	3.92	0.53	.59	0.29	3.53	.000
Negative Mood	1.99	0.77	.80	1.74	0.75	.81	0.38	4.82	.000
Positive Mood	3.25	0.81	.82	3.40	0.89	.80	0.52	3.01	.003

Note: Table entries were computed for participants with stress and mood data for the pre- and post-training sessions. The sample size was n = 277 except for negative mood (n = 265) and positive mood (n = 267). Alpha is Cronbach's internal consistency estimate of reliability. "r_{tt}" is the correlation of the pre-training scores with the post-training scores (Nunnally & Bernstein, 1994, pp. 233-236).

Firefighting training decreased stress. Role clarity increased during training from a pre-training average rating very close to the "Neither Agree Nor Disagree" anchor of the rating scale (i.e., 3) to near the "Agree" anchor (i.e., 4). Role conflict dropped from an anticipatory mean value of "Neither Agree Nor Disagree" to a post-training rating about one-third of the way from the original rating and "Disagree" (i.e., 2).

b. Self-Efficacy

Self-efficacy measures the person's perception that he or she can master the tasks required in firefighting training. Self-efficacy scores also include the perception that the tasks can be performed with relative ease and/or with simple persistence.

Eight items in the Stress Profile questionnaire were designed to measure self-efficacy. Principal component analysis indicated the presence of a small second factor in these items, but the total set of items was used to measure the construct. This procedure corresponded to the concept of generalized self-efficacy that was the basis for the development of the self-efficacy items. The resulting scale had a rather low, but marginally acceptable, internal consistency estimate of reliability (cf. Table 18).

Average self-efficacy increased, but the gain was modest. The initial rating was approximately three-fourths of the distance from "Neither Agree Nor Disagree" to "Agree." The final rating was approximately nine-tenths of the distance. The absolute average gain (0.14 rating points) translated to an effect

40

size of 0.25 using the pre-training standard deviation (SD) as the frame of reference. This change represents a small effect (Cohen, 1988) in the predicted direction.

All of the differences between anticipatory stress and post-training stress were comparable for females and males. None of the gender by time interactions was statistically significant ($p > .117$ for each).

c. Positive and Negative Moods

Mood assessment consisted of 12 items, six selected to measure negative mood and six selected to measure positive mood. Extensive prior research indicates that these two general dimensions provide a useful summary of mood (Watson & Tellegen, 1985). Vickers and Hervig (1987) showed that this model, applied to military personnel, was not merely a response style artifact; it extracted the valid variance from mood reports in this population.

Given the selection of items, it was anticipated that the mood items would produce two clear dimensions. Principal components analysis produced only two eigenvalues > 1.00 in both the pre-treatment and post-treatment data (Pre-treatment eigenvalues = 4.13, 2.29, 0.90; Post-treatment eigenvalues = 3.94, 2.72, 0.88).

The component loadings confirmed the expected mood structure (Table 19). The six negative mood items defined a single dimension in each analysis; The six positive mood items defined the other dimension. Based on this evidence and the prior studies of mood structure, the items were combined to form the *a priori* scales planned for positive mood and negative mood. The resulting six item scales had acceptable reliability (cf. Table 3).

Table 19. Principal Components Structure of Affective Distress

Item		Component 1		Component 2	
		Pre	Post	Pre	Post
37	Sad	.78	.83	−.07	−.07
33	Blue	.74	.70	−.18	−.11
35	Angry	.73	.78	−.06	.00
38	Annoyed	.68	.82	−.09	.02
32	Uneasy	.68	.58	−.02	.02
29	Depressed	.66	.68	−.23	−.26
39	Happy	−.17	−.17	.79	.80
40	Energetic	−.16	−.14	.79	.79
31	Cheerful	−.12	−.07	.79	.77
36	Active	−.16	.04	.70	.74
34	Satisfied	−.24	−.19	.60	.68
30	Vigorous	.27	.13	.56	.57

Note: The number in front of each adjective indicates the item number within the Stress Profile questionnaire.

Moods were positive overall prior to firefighting training, but improved further during firefighting training. The assertion that moods initially were positive overall is based on the observation that the average rating for positive mood was slightly above the "Neither Agree Nor Disagree" rating (3.25), while the average rating for negative mood was close to the "Disagree" anchor point (1.99).

The inference that mood improved during firefighting training is based primarily on a drop in the level of negative mood. The average rating for negative mood dropped one quarter of a point to a value one-fourth of the way between "Disagree" and "Strongly Disagree." Using the pre-training mood as the reference, this change amounted to a small effect size (ES = 0.32).

The inference that mood improved was supported by a marginal trend toward increased positive mood. The average positive mood rating increased by approximately one-seventh of a rating point (0.15). Whether this increase in positive mood should be regarded as a useful indicator of change during firefighting training is uncertain. The change was statistically significant ($p < .003$), but the effect size was smaller than that for negative mood (ES = 0.19 vs. ES = 0.32). The positive mood effect size fell just below Cohen's (1988) recommended minimum for an important effect (i.e., ES = 0.20).

3. Evaluation of Pre-Training Status (Task 5)

a. Randomization Effects

Study participants were randomly assigned to treatment groups. While the procedures followed should have produced comparable groups, differences might have arisen by chance. For this reason, analyses would benefit from the introduction of appropriate controls for those variables that were unevenly distributed across the treatment conditions.

A one-way analysis of variance was conducted to verify that the experimental groups were comparable. The analysis compared the three treatments. Measures that described the individuals prior to treatment were the dependent variables.

The treatment groups did not differ significantly on any pre-treatment variable. This statement was true for Armed Forces Qualification Test (AFQT) ($F_{2,274} = 0.17$, $p < .849$) anticipated role clarity ($F_{2,289} = 0.98$, $p < .377$), anticipated role conflict ($F_{2,289} = 1.54$, $p < .216$), and anticipated self-efficacy ($F_{2,289} = 0.86$, $p < .427$). AFQT is a general indicator of intelligence. Similarly, the treatment groups reported comparable initial levels of affective distress, including both negative affect ($F_{2,280} = 1.40$, $p < .249$) and positive affect ($F_{2,280} = 0.54$, $p < .583$). Clearly, the treatment groups were well matched with respect to mental ability, anticipated stress, and initial affective stress status.

b. Gender Differences in Ability

Females (mean = 61.54, SD = 19.54) scored higher on the AFQT than males (mean = 55.94, SD = 20.76). This difference was only marginally significant statistically ($t = 1.93$, 274 df, $p < .056$) using a two-tailed significance test. A two-tailed test was appropriate because no *a priori* prediction had been made regarding the direction of gender differences in mental ability.

c. Gender Differences in Anticipated Stress

The stress measurements taken prior to firefighting training are best regarded as measures of anticipated stress. These measures asked recruits to indicate what they believed the firefighting training would be like. This perception presumably was based on the recruits' prior experiences in basic training, prior

beliefs about the nature of the course, and so on. These initial stress perceptions obviously were not reports of any actual experience in firefighting training. For these reasons, it was desirable that the initial ratings be differentiated from the actual stress experienced during the training. The label "anticipatory" was applied to provide this differentiation.

Females and males had comparable anticipations about stress in firefighting training (Table 20). Differences between the two groups were not statistically significant for role ambiguity, role conflict, self-efficacy, and negative mood. Females reported slightly higher levels of positive mood, but this trend was only marginally significant using a two-tailed significance test. A two-tailed significance test for the difference in positive mood scores was appropriate given the observed direction of the difference. The initial expectation was that females would have higher stress levels (i.e., less positive mood in the case of this indicator).

Table 20. Pre-Training Gender Differences

Scale	Females Mean	SD	Males Mean	SD	t	Sig.
Role Ambiguity	2.72	0.68	2.65	0.65	0.79	.433
Role Conflict	2.95	0.66	3.02	0.60	0.83	.405
Self-Efficacy	3.79	0.49	3.77	0.57	−0.33	.744
Negative Mood	1.94	0.68	1.76	0.77	0.88	.381
Positive Mood	3.42	0.77	3.20	0.82	−1.95	.052

4. Evaluation of Post-Training Status (Task 5)

a. Academic Performance

The average female (mean = 3.94, SD = 0.44) scored lower on the examination than the average male (mean = 4.15, SD = 0.45). The difference was statistically significant ($t = 3.56$, 296 df, $p < .001$). The lower average score for females was obtained even though they had higher than average AFQT scores. Because higher AFQT scores generally were associated with higher examination scores ($r = .400$, $t = 7.23$, $p < .001$), females' higher average AFQT scores led to the expectation of higher than average performance on the firefighting examination.

A regression equation was developed to explore female-male differences in examination performance adjusting for the difference in general mental ability. The equation was

$$T' = 3.613 + (.00871 * AFQT)$$

where T' is the predicted test score.

An adjusted test score (T_{adj}) was computed to measure performance controlling for ability. This score was computed by subtracting the predicted test score from the observed test score (i.e., $T_{adj} = T - T'$ where T is the test score). The adjusted scores were positive if performance was higher than expected given the person's ability level, and negative if performance was lower than expected.

43

b. Treatment Effects on T_{adj}

Treatment effects on academic performance were evaluated by comparing the T_{adj} scores of females and males within each treatment group. The comparison of females and males in the control group supported the initial claim that females perform relatively poorly in this component of basic training. The average score for females (mean = $-.226$) was lower than that for males (.085) as expected. Absent any treatment, therefore, the female-male difference in test performance was d = $-.311$; this difference was statistically significant ($t = -3.24$, 86 df, $p < .001$, one-tailed).

The RM treatment did not improve females' performance on the firefighting examination. Females who received this treatment still scored below average after adjusting for their general mental ability (mean = $-.240$). Males who received this treatment had scores that were, on the average, virtually identical to their predicted scores (mean = .019). The difference between females and males ($d = -.259$) was only slightly smaller than that under the control conditions ($d = -.311$). Females and males still differed significantly on performance in this treatment ($t = -2.64$, 88 df, $p < .005$, one-tailed).

The AO treatment improved females' performance on the firefighting examination. Females in this treatment scored only slightly below the level expected based on their AFQT scores (mean = $-.074$). This deficit was only one-third to one-fourth that found in the control group and the recruits receiving the RM treatment. Males who received the AO treatment scored slightly higher (mean = .095) than expected based on their AFQT scores. The difference between females and males, therefore, remained moderately large ($d = .169$), but was reduced in size relative to that in other treatments. The remaining female-male difference was not statistically significant ($t = 1.61$, 85 df, $p < .056$).

The AO treatment effect is placed in perspective by several comparisons. To begin with, the female-male difference in this treatment was only about one-half that in the control group (d = $-.169$ vs d = $-.311$). The trend toward a smaller gender difference for the AO group was consistent with the hypothesis that this treatment would improve female performance relative to male performance. However, although the trend was in the predicted direction, it was only marginally significant (difference = .142, t = 1.34, $p < .092$, one-tailed).

Comparing the two treatments, each treatment reduced the performance deficit in females, but the AO treatment was more effective. The AO treatment cut the performance deficit for female recruits by 46% compared to a reduction of only 17% for RM. Given the small sample sizes in the present study, the AO gain was only marginally significant, but the RM gain did not even approach significance.

c. Treatment Effects on Stress

Treatment effects on stress were estimated using anticipated stress as a covariate. This procedure was used instead of a repeated measures analysis. The repeated measures analysis assumes that the stress scales measured the identical construct at both the pre- and post-training administrations. This assumption was questionable on logical grounds for the stress and efficacy scales. In these instances, the post-training scores reflect perceptions and feelings based on actual experiences in firefighting training. Recruits cannot reasonably be expected to know accurately what those experiences will be prior to training. Pre-test variables, therefore, must reflect generalization from other parts of basic training, generalized tendencies to perceive stress, trait elements of mood, or other similar factors. Arguably, the measurements taken at the two times represent distinct constructs (e.g., capacity for stress vs. actual stress), in which case a difference score such as that which would be evaluated in a repeated measures design is difficult to interpret. Using the pre-training scores as covariates of the post-training scores can be seen as a method of adjusting for the continuing influence of those factors that determined the individual differences in

anticipation. Removing the effects of those factors should leave a purer indication of the effects of training on stress.

Mood scores presumably measured the same constructs at both testing times. These scales, therefore, could reasonably have been evaluated using a repeated measures analysis of variance. The analysis of covariance was employed for these distress scales to make those analyses comparable to the analyses for the other stress indices.

Table 21 summarizes the findings for stress as a function of gender and treatment group controlling for anticipatory stress. The only significant main effects were those obtained for gender and treatment on role conflict. These effects were present because females experienced less conflict than males, and the RM treatment was associated with less conflict. The estimated mean score adjusted for pre-training anticipatory conflict was 2.50 (SE = .08) for females compared to 2.74 (SE = .05) for males. The corresponding estimates for the treatments indicated that the RM treatment had the lowest value (mean = 2.52, SE = .08) followed by the control group (mean = 2.65, SE = .09) and the AO treatment group (mean = 2.69, SE = .08). The primary difference between treatment groups appeared to be the contrast between the extreme groups. The 95% confidence intervals for the RM treatment group (2.36, 2.68) did not include the mean for the AO group; the 95% confidence interval for the AO treatment group (2.54, 2.85) did not include the mean for the RM treatment group.

Table 21. Treatment Effects on Stress Indicators

Scale	Gender		Treatment		Gender x Treatment	
	F	Sig.	F	Sig.	F	Sig.
Role Ambiguity	1.50	.22	0.03	.973	0.26	.774
Role Conflict	11.56	.001	6.45	.012	1.22	.298
Self-Efficacy	0.37	.542	2.07	.128	2.35	.097
Negative Mood	0.47	.495	0.72	.487	1.55	.215
Positive Mood	0.57	.451	0.66	.518	0.30	.738

d. Extended Analysis of Interactions

Table 21 also indicated the presence of one marginally significant interaction between gender and treatment. This marginal effect was examined in more detail because the interaction was a diffuse significance test (i.e., a test involving more than one degree of freedom; cf. Rosenthal & Rosnow, 1984). In this case, one or more meaningful differences might be concealed because they were mixed with a number of other trivial differences. Given that this phase of the study was only an initial test of the treatments, exploration of this possibility was reasonable.

The interactions between gender and treatment approached significance in the omnibus analyses for the adjusted examination score and self-efficacy. Examination of the average scores in the different treatments suggested that this circumstance occurred because the treatments had very specific effects on females. Despite the weakness of the trends for these differences, further evaluation was appropriate given the preliminary character of the study. The results of the evaluation will direct future refinements of the treatments.

Self-Efficacy

Examination of the mean scores for self-efficacy suggested that the RM treatment was the primary source of the treatment x time interaction for this variable. In general, females and males had comparable levels of anticipated efficacy in firefighting training. Post-training efficacy generally was comparable for females and males adjusting for anticipated efficacy. However, females who received the RM treatment showed exceptionally high self-efficacy after firefighting training.

An adjusted self-efficacy post-training score was computed. Self-efficacy was prior to firefighting training and after firefighting training. The difference between the predicted and observed post-firefighting training scores provided the adjusted post-training score. Given that anticipated self-efficacy quite probably reflected a generalized self-efficacy based on general life success and success to that point in basic training, the adjusted post-firefighting training self-efficacy score can be interpreted as an index of situation-specific self-efficacy scored as a deviation from generalized self-efficacy.

Combining the AO and control groups into a single group, recruits receiving the RM condition could be compared to all others. The resulting 2 x 2 ANOVA indicated that the RM treatment had a significant main effect ($F_{1,273} = 4.18, p < .042$), but there also was an interaction of gender with treatment status. The average adjusted scores indicated that treatment status had little average effect for males (RM = .01; Others = .02). Among females, the RM treatment was associated with higher adjusted self-efficacy (0.18). Other treatments led to lower self-efficacy (−0.13). The difference between the two groups of females was statistically significant ($t = 2.24, p < .028$) and translated to a moderate effect size (ES = 0.50).

Negative Mood

The existence of a significant effect of the RM treatment on one psychological stress indicator motivated a search for other indicators with a similar pattern. Examination of the average pre-training and post-training scores also suggested that the same pattern contrasting the RM treatment with the AO treatment and controls was present for negative mood. In this case, the interaction effect for the original 2 x 3 ANOVA was not even marginally significant ($p < .215$; cf. Table 21), and even the interaction for the 2 x 2 ANOVA was only marginally significant ($F_{1,261} = 2.78, p < .097$).

The negative mood differences were of interest despite their weak statistical significance because they paralleled the self-efficacy differences. Once again, the RM treatment had little effect on males (RM = .07; Other = −.01) and a positive effect on females (RM mean = −.24; Others mean = .04). The difference between females receiving the RM treatment and other females was statistically significant ($t = 1.76$, 67 df, $p < .042$, one-tailed). A moderate effect size (ES = .43) was evident.

Examination Performance

Group differences in examination performance were explored in more detail using T_{adj} as the criterion measure. The group means from the gender x treatment analysis indicated that the AO improved this outcome for females. Females who received the AO treatment had examination scores that were only slightly less than the score that would be predicted based on their AFQT scores (Avg. = −.07, SD = .39). The hypothesis that the true mean score for females receiving this treatment is zero (i.e., that the females receiving this treatment performed up to the level expected based on their mental ability) cannot be rejected ($t = −0.82$, 20 df, $p < .211$, one-tailed). Females who did not receive the AO treatment scored below their predicted performance (mean = −.23, SD = .37, n = 40). This deficit was significantly below zero ($t = 3.93$, 39 df, $p < .001$). The performance deficit in these females was lower than that for females who received the AO treatment, but the difference was only marginally significant ($t = 1.58, p < .061$).

46

Comparing females to males who received the same treatment also illustrates the potential value of the AO for improving females' academic performance. Among those receiving no treatment (i.e., the control condition), males performed significantly better than females (male mean = .09; female mean = −0.23; t = 3.24, 86 df, $p < .001$, one-tailed). The same trend was evident in the RM treatment (males = 0.02; females = −0.24; t = 2.64, 88 df, $p < .005$, one-tailed). The female-male difference was substantially smaller and only marginally significant statistically in the AO condition (males = 0.10; females = −0.07; t = 1.61, $p < .056$).

Expressing the impact of the AO on academic performance as an effect size provides another perspective on the effectiveness of this treatment. On the average, the adjusted score for females in the control and RM groups was 0.29 points lower than that for males in the same treatments. The comparable difference was 0.17 between female and male recruits receiving the AO treatment. Stated this way, the AO provided an improvement of 0.12 points relative to males receiving the same treatment. This gain represents an effect size of ES = 0.30 using the pooled SD for all males as the denominator for the computation. This ES value is near the middle of Cohen's (1988) small effect size range.

5. Individual Differences in Stress and Performance (Task 5)

The analyses described thus far examined group differences in stress and performance. Those analyses compared average scores for different groups, and treated individual variation within groups as error variance. This approach did not test for possible associations between individual differences in stress and individual differences in performance. Given that it is individuals who actually perceived stress and performed on the examination, the grouping process may have been somewhat misleading. Additional analyses, therefore, were undertaken to examine associations between stress and performance with the individual as the unit of observation.

The additional analyses evaluated the hypothesis that stress affects performance by T_{adj} to adjusted stress scores. The adjusted scores corrected post-training stress and distress reports for the effects of pre-training anticipatory stress and distress. Table 22 shows the results of these analyses for the full sample, all females, all males, all participants in the control group, all participants in the AO treatment, and all participants in the RM treatment.

a. Full Sample

Anticipatory stress was not related to T_{adj} (r = .007 to r = .066). Adjusted post-training scores generally were unrelated to T_{adj}, but good performance was associated with a larger than expected decrease in negative mood ($r = -.14, p < .024$).

47

Table 22. Adjusted Stress Scores as Predictors of Adjusted Academic Performance

Scale	Total Sample	Females	Males	Control	Advanced Organizer	Role Model
Role Ambiguity	.006	.035	.011	−.023	.037	.008
Role Conflict	−.023	.178	−.134	−.047	.001	−.053
Self-Efficacy	−.007	−.136	.032	.117	.021	−.156
Negative Mood	−.144*	.016	−.208**	−.323**	−.094	−.242*
Positive Mood	−.067	.149	−.108	−.046	−.130	.008

*$p < .05$ **$p < .01$

Note: Sample sizes varied slightly depending on the amount of missing data for the different measures. The ranges of sample sizes were: Total Sample = 247–259; Females = 58–61; Males = 189–198; Control = 82–86; AO = 82–85; RM = 83–88.

b. Gender

It was thought that gender might affect the associations between stress and performance. However, gender comparisons produced ambiguous results. First, the T_{adj}–negative mood association was limited to males ($r = -.21$). This association was essentially zero among females ($r = .02$). However, the difference between the correlations was only marginally significant statistically ($z_{diff} = 1.48$, $p < .070$) using the Fisher r-to-z transformation (Hays, 1963, p. 532). This result was ambiguous because a simple "Is the association significant in both groups?" test for differences would lead to the inference that the groups differed while asking "Are the two correlations significantly different?" which would lead to the inference that the females and males did not differ.

The results for a second correlation were ambiguous because they demonstrated the opposite pattern. Among females, adjusted role conflict was positively related to academic performance ($r = .178$). Among males, the correlation was negative ($r = -.134$). The *difference* between these correlations was statistically significant ($z_{diff} = 2.20$, $p < .014$), but neither of the individual correlations was significant. Thus, the data indicated a significant difference from one perspective while making it impossible to rule out the following equality: $r_{females} = r_{males} = .00$.

The fact that the determination of statistical significance is dependent on the establishment of a significance criterion adds to the ambiguity of the role conflict–T_{adj} correlations for females and males. The observed difference was labeled "significant" using the common $p < .05$ criterion. However, role conflict was one of five potential predictors of academic performance considered in the analysis. Conservative significance testing would apply a Bonferroni adjustment to maintain the experiment-wide error at $p < .05$ (Harris, 1985, pp. 7-9). In this case, the significance criterion would move to $p < .01$; the observed result would be classified as nonsignificant using that criterion. Given these additional considerations, the evidence for female-male differences in stress-test performance relationships is equivocal.

48

c. Treatment

Treatments also could affect stress-performance relationships. This possibility received some support from comparisons for negative mood. Adjusted negative mood was significantly related to adjusted academic performance in the control ($r = -.323$, n = 82, $p < .003$, two-tailed) and RM ($r = -.242$, n = 83, $p < .028$, two-tailed) treatment groups, but not the AO group ($r = -.094$, n = 82, $p < .401$). The differences appear to be substantial, but they are not statistically significant. Comparing the AO treatment to the control group, the difference approached significance ($z_{diff} = 1.51$, $p < .132$, two-tailed). The difference between the AO treatment and the RM treatment did not even approach significance ($z_{diff} = 0.96$, $p < .338$, two-tailed). Two-tailed significance tests were reasonable in these comparisons because there were no *a priori* hypotheses about the location and direction of the differences tested. These results were ambiguous in the same way that the gender difference in the negative mood–T_{adj} relationship was ambiguous.

6. Conclusions and Recommendations from Year Two (Task 6)

The initial evaluation of the new instructional technologies supports several general conclusions. The results confirm that females experience more difficulty than males in learning the classroom material in firefighting training. The results indicated that the AO treatment reduced the relative deficit in academic performance, but the RM treatment did not. In contrast, the results indicated that the RM treatment had some positive effects on distress and efficacy, but the AO treatment did not.

The evidence provided no reason to believe that stress was a mediator of academic performance, but no inference could be drawn about stress and leadership or team performance because suitable criteria were not available. Finally, preliminary descriptions of the patterns of individual interactions with the AO treatment were identified that may provide a basis for improving the treatment. Each of these points is considered in more detail in the following paragraphs.

a. Performance of Females

Females did relatively poorly on the academic test. The control group provides the best estimate of this deficiency as it currently exists in training because these females essentially received the same treatment that recruits in general receive at present. The average adjusted score for the females in the control group was −.23. This figure was 0.58 standard deviation below the average adjusted score for males. The difference represents a moderately large effect size in behavioral research (Cohen, 1988). This evidence quantifies the size of the problem being addressed with greater precision than was possible at the study outset.

b. Effects of Treatment

AO Treatment Effect

The AO treatment improved the academic performance of female recruits. A slight performance decrement relative to ability still was evident in this group (mean = −.07), but the deficiency was less than one-third the size of that noted in both the control and RM treatment groups. Furthermore, statistical tests indicated that the observed AO deficiency was not significantly different from zero or significantly different from that of males going through the same treatment. The overall pattern of findings indicates that the AO treatment reduces the female academic performance deficiency and may eliminate it altogether.

RM Treatment Effect

The RM treatment did not affect academic performance, but this result was not surprising. The RM treatment was designed to improve performance in the hands-on portion of firefighting training. The emphasis on teamwork and leadership was expected to be rewarding when recruits had to communicate and work together to put out fires. The project was undertaken with the belief that suitable hands-on firefighting measures would be available to assess individual differences in performance. These measures were not developed. The improved sense of efficacy and better mood at the end of training suggest that female recruits who went through the RM treatment felt they did better. Recruits' spontaneous comments also suggested they felt this was a useful treatment. Thus, it seems likely that RM affected performance, but not in the areas that could be directly measured.

Hawthorne and Placebo Effects

The fact that the RM and AO treatments affected different outcomes is important. The AO treatment affected academic performance, while the RM treatment affected stress. This difference helps rule out a Hawthorne effect (Shaughnessy & Zechmeister, 1985, pp. 336-338) or placebo effect as the explanation for the treatment effects. A Hawthorne effect would occur if simply giving some people more attention resulted in improved performance. A placebo effect would occur if a treatment effect was present only because the recruits receiving the treatment believed it would affect them. In the present case, there was no reason for recruits to believe that one treatment would be useful, but the other would not. Given these assumptions, both groups received comparable attention and received treatments that they believed would be efficacious. Neither assertion applied to the control group. The result is that the placebo and Hawthorne effect interpretations both would differ from the control group, but would be comparable to one another. A pattern of differential effects for the AO and RM treatments is contrary to these expectations.

Differential Treatment Effects

The differential effects for the AO and RM treatments become more plausible when it is noted that these effects occur where they might be expected to occur, based on the nature of the treatments. The AO treatment was designed to prepare a cognitive basis for learning academic materials. The RM treatment illustrates patterns of interpersonal interaction and behaviors that are basic to role communication and role enactment and, therefore, are basic to the occurrence of role ambiguity and role conflict. The AO treatment effect may be the result of changes in the cognitive processes involved in learning the academic material in firefighting. The RM treatment may modify the stress perceptions and other psychological reactions to firefighting training that develop out of the teamwork component of training. Taken together, these considerations help rule out nonspecific causal explanations (e.g., Hawthorne and placebo effects) in favor of causal explanations based on the specific content of the treatments. The pattern of results, therefore, provides a basis for inferring that the treatments did in fact cause the differential performance and stress effects observed in this study.

Stress Effect

Stress reduction was not associated with improved academic performance. This inference is based on the lack of connection between stress and performance. First, the AO treatment improved females' academic performance, but did not reduce stress. Second, the RM treatment reduced some stresses, but did not improve academic performance. Third, individual differences in stress were not related to individual differences in performance. The only suggestion of a relationship was that negative mood and T_{adj} were significantly related. However, that association could be explained as well by assuming that poor

academic performance led to increased depression, anger, and so on. Also, even the association between negative mood and T_{adj} was absent in the one treatment that actually modified academic performance (i.e., the AO treatment). Exploratory analyses suggested that the relationships between stress and academic performance might differ for males and females or between different treatment groups, but chance sampling variation could not be completely ruled out as a basis for those differences.

The RM treatment reduced stress among females. Females receiving the RM treatment showed gains in self-efficacy and reduced negative mood at the end of firefighting training. Because males and females were comparable on these two measures at the beginning of firefighting training, these changes mean that females in the RM treatment stood out as being lower on these stress indicators relative to all other males and females.

Females and males generally reported comparable stress levels. The only significant difference between genders was for females to report lower role conflict after firefighting training. Given evidence that females possess personality characteristics associated with reporting more negative affect (Feingold, 1994), females might have been expected to have higher negative mood and perhaps lower positive mood. However, in this setting, the difference was not evident.

The finding that females and males experience comparable stress levels is consistent with prior research (Martocchio & O'Leary, 1989). Other reviewers have suggested that females report more affective stress than males (Jick & Mitz, 1985). However, the effect sizes involved were not reported in that review, and the sample sizes for the majority of the studies were quite large (n > 700). This evidence may point to a general tendency toward higher affective stress in females, but the difference could be trivial. Furthermore, the recruit situation is different from the general population in that males and females have the same roles, face the same living situations, etc. Even if female-male differences in affective stress are present in the general population, the differences might disappear if the two groups were equated for occupational status, education, family situation, and other factors as they are, to a great extent, in basic training. A general belief that females are more susceptible to stress than males may exist. If so, based on the results of this study, that belief appears to be a misleading stereotype.

c. Limitations of Year Two Study

Sample size and contamination were two main limitations in the Year Two study. Research during Year Three of the study addressed these limitations.

Sample Size

Initial findings are based on small sample sizes that provided the intended preliminary evaluation of the treatments. The analyses frequently explored marginally-significant trends in the data without allowing for the extent to which these explorations could capitalize on chance. This approach was appropriate for a preliminary evaluation that was intended to provide information useful for refining the study hypotheses. Strict adherence to traditional significance testing criteria combined with broad general hypotheses might have resulted in important effects being overlooked by using diffuse significance tests. However, the approach taken here can be interpreted only as establishing a set of refined working hypotheses for further study. The findings need to be replicated with larger sample sizes and drawing on recruits who go through training at different times of the year. The latter requirement arises because important recruit qualities change over the course of the year. The effectiveness of the treatments could depend on one or more attributes of recruits. It is impossible to say *a priori* whether the recruit differences will increase or decrease the effectiveness of the treatments.

The sample size in the initial study was also too small to study the effects of the individual differences in learning style. Such differences may impact the effectiveness of the AO treatment. Preliminary evidence was gathered during Year Two that indicates the existence of individual differences in how people interact with the computer in this condition (e.g., choosing the linear approach versus the concept map, the number of questions asked, the types of questions asked, etc.). These differences may influence how well the recruits learn firefighting material. If so, studying the effects of the differences on performance could provide helpful clues for refining the AO treatment to improve its overall effectiveness. General principles relevant to all types of technical training might be identified, but it is too early to tell. This problem will be overcome by increasing the sample size, quantifying differences in the pattern of interaction with the computer, and relating those differences to performance outcomes.

Contamination

Some contamination may have occurred in the original study design because of shared interest on the part of recruits and instructors. "*Contamination* occurs when there is communication of information about the experiment between groups of subjects" (Shaughnessy & Zechmeister, 1985, p. 335, italics in the original). Each division studied included recruits who were exposed to the AO, RM, and control treatments. Recruits in the control group knew that others were receiving the experimental treatments. Instructors were concerned that materials learned in the treatments would help only some of the recruits in their division when all of the recruits could potentially benefit. Anecdotal evidence indicates that exposure to the treatments stimulated recruits to change their behaviors in several ways, including asking more questions in later parts of classroom training and participating more actively in discussions. These behaviors changed the nature of subsequent learning opportunities not only for the recruits within a particular treatment, but also for all the other recruits in their division. This type of contamination may have acted to influence treatment effectiveness. The potential contamination problems will be minimized during next year's data collection by changing the research design so that a single treatment will be administered to all recruits within a division or within paired divisions.

d. Recommendations

There were four main recommendations from the Year Two study, as described in the following paragraphs.

- **Focus on AO Treatment**

    The AO treatment would be the primary focus of the Year Three work. This treatment demonstrated a sizable effect on academic performance, a meaningful criterion for training personnel. That demonstration needs to be replicated. Preliminary evidence was obtained indicating that different people interact differently with this treatment. The small sample size makes it impossible to say whether differences in the interaction pattern affect the outcome from exposure to this treatment. The sample size needs to be increased for this treatment in order to permit more detailed analysis of the relationship between patterns of interaction and treatment effectiveness.

    Replication of the RM treatment will not be implemented. While the RM treatment appeared to be effective, the effects on behavioral variables cannot be demonstrated in the current firefighting training context. For this reason, it is not possible to directly demonstrate that the treatment has a positive effect on outcomes that are central objectives of basic training. Reduced stress may be good, but it is valuable in a training context primarily if it demonstrates a positive effect in achieving learning objectives. The stress effects provide reason to believe that the RM treatment can affect training outcome criteria, but sound behavioral assessments of leadership and/or team effectiveness

would be required to extend this work. Such measures could be developed, but even then the highly structured, carefully controlled setting of firefighting training might be a poor place to examine these effects. Recruit comments suggest that other situations, such as Service Week, provide more extensive leadership and team performance opportunities. Some recruits commented that they felt this treatment would have helped them during that time period. However, developing this line of inquiry is beyond the scope of available resources for this project. The RM treatment should be given consideration for further research if more resources become available, but until then, the benefit seems greater for focusing on the AO treatment.

- **Reduce Contamination**

The research design will be modified to eliminate possible sources of treatment contamination. The initial design was guided, in part, by a desire to gather useful information on both AO and RM treatments with as little intrusion into training as possible. Based on the initial findings, we will implement a revised design that exposes all of the recruits in a division to a single treatment. The control group will be comprised of recruits from divisions who go through the regular training schedule with no treatment.

- **Evaluate Individual Differences in Learning Style**

Research during Year Three of the study will capitalize on the opportunity to examine individual differences in learning style. Demographic, academic and educational variables will be assessed for their role on how academic data is assimilated and how it impacts the curricular successes of male and female recruits. Students interact with the AO treatment in a manner that is consistent with their individual learning styles. AO interaction is defined as selection of method (concept map or linear) and selection of questions within the AO treatment. By evaluating individual differences in learning styles, we will gain insight into which strategies are the most successful at preparing students of different genders, ethnicity, education, intelligence, and age for the academic portion of the course.

Two more data collections will be conducted. The first will provide information to better quantify individual differences in patterns of interaction with the AO. This data will increase the sample size sufficiently to estimate the influence of these patterns on academic proficiency. The second study will replicate those findings and might involve some modifications to the AO treatment itself. The timing of the two data collections will be chosen to verify that the AO effect is present for recruits who enter the service at different times of year.

- **Evaluate Gender Differences in Stress-Performance Associations**

Stress assessments will be retained as part of the first data collection effort in Year Three. Although the overall evidence from this study suggests that stress and performance were not generally related, the comparison between the AO and control groups appears to be the most likely place to find an exception to this generalization. Among controls, poor performance was associated with more negative mood at the end of firefighting training. Receiving the AO treatment appeared to reduce this effect, but the trend was only marginally significant. Replication of the initial finding is needed to determine whether this difference indicates an effect of AO. Also, the evidence suggested that females and males differ with regard to stress-performance relationships. Perhaps this difference exists primarily in specific treatment conditions. The sample sizes in the present study were too small to evaluate gender differences in stress-performance associations within specific treatments. Increasing the sample size for the control and AO groups will permit more detailed analysis of this possibility.

53

C. YEAR THREE

1. Experimental Methods from First Data Collection Effort, Year Three (Tasks 4 and 5)

During Year Three of the project, the research team completed tasks 4, 5, and 6 of the approved Statement of Work.

a. Procedures for Implementation of Intervention and Data Collection

One "integrated" division per day received the AO treatment. An integrated division consisted of one division of male recruits and one division of female recruits, integrated and processed through basic training as a single division. The procedures for implementation of the treatment and data collection were as follows.

In the morning, the FAST lab was prepared. The computers were logged onto and the AO treatment was waiting at the sign-on screen for the recruit. When the assigned division for that day arrived, they were brought into the lab. The recruits had already signed the consent forms, which were administered by the Navy. The recruits were introduced to the topic of the study.

In order to accommodate the number of computers available in the lab and the recruits, 40 recruits at a time received the treatment. Upon entering the lab, the recruits were given specific instructions pertaining directly to the AO treatment. When the recruits completed the treatment, they were released back to their RDC. The recruits completed the firefighting course at which time they took the firefighting academic test.

The recruits in the control group signed the consent forms and received the conventional firefighting training, which included taking the academic firefighting test.

b. Residualized Stress and Performance Scores

Residualized change scores for stress and performance measures were the dependent variables in the ANOVAs. First, post-training stress were regressed on the scores for the same variable obtained prior to firefighting training. The regression equation was used to predict the level of post-training stress for each participant based on his or her pre-training score. The predicted score was subtracted from the observed post-training score. The difference was the residualized score.

Residualized gain scores were used to estimate change in preference to simple difference scores (i.e., post-training score - pre-training score). Residualized scores were favored because the stress scales arguably did not measure the same construct before and after training. Anticipations of stress based on past basic training experiences and any hints about the nature of firefighting training that the recruits may have heard are not equivalent to the actual experience. Computing a difference based on two scores that reflected different constructs would be the statistical equivalent of comparing apples and oranges. The residualized gain approach requires only that the two variables be empirically related, not that they measure the identical construct.

Residualized gain scores also were preferred because they provided a consistent method of quantifying the hypothesized effects of the AO. Residualized scores were the only reasonable option for the academic performance because subtracting Armed Forces Qualifying Test (AFQT) scores from academic test scores does not yield a meaningful estimate of performance.

The computation of the residualized academic test scores was slightly different than those for stress. Predicted test performance was not based on a pretest. Instead, the academic test residual was based on the regression of test scores on AFQT scores. The AFQT is a measure of general intelligence that predicts performance in a wide range of military training settings. The residualized test score was the difference between the test score the person obtained and his or her predicted test score based on his or her AFQT score.

Residualized scores must be properly interpreted. A person whose post-training score is exactly equal to what would be predicted given his pre-training stress scale score or AFQT will have a score of zero. A negative residual indicates that the person had a lower posttest stress or academic test score than expected based on his or her pre-test score or AFQT. A positive residual indicates a higher than expected score.

Note that the average score on a stress indicator is not part of the residual score. For example, in this study, the average score on Role Clarity increased by 0.49 points from pre-firefighting training to post-firefighting training. Despite this change, the average Role Clarity residual for the full sample will be zero. The effects of experimental treatments will be indicated by deviations from zero. The residualization process removed the average gain from the individual scores. The residualized gain scores accurately reflect individual differences in change and differences between treatments.

c. Analysis Procedures

Bivariate Relationships. Bivariate relationships were examined using four different analysis procedures.

1. Independent Sample t-tests. These tests were performed when two specific groups in the sample were compared on stress and performance measures. For example, this procedure was used to test for gender differences in anticipated stress. Effect size (ES) was expressed as a point biserial correlation where appropriate in these analyses.

2. Paired sample t-tests. These tests assessed differences between the anticipated stress measured at the beginning of firefighting training and the actual stress perceived at the end of firefighting training.

3. One-way analysis of variance (ANOVA). This analysis technique was employed to evaluate relationships between stress and performance in relation to categorical measures that had more than two groups. For example, one-way ANOVA was used to assess the relationship between age and the stress and performance measures. The eta-squared (ε^2) value for the differences was the measure of effect size in these analyses (Hays, 1963).

4. Analysis of Cross-Tabulations. Associations between pairs of categorical variables (e.g., gender, learning style, ethnicity) were evaluated by χ^2 tests. These tests were based on two-way tables defined by the variables of interest (e.g., gender and a learning style variable). Significant associations were followed by descriptions of the group differences in proportions to describe the association.

5. Pearson Product-Moment Correlations. These correlations were computed when the bivariate relationship of interest was the association between two conceptually continuous variables (e.g., stress and performance).

Multivariate Models. Multivariate models refer to analyses that explored the joint effects of two or more predictor variables on the stress and performance measures. The analyses conducted for this purpose included:

1. Multiple Regression. Standard linear multiple regression procedures were used to predict academic test performance based on AFQT scores and AO Test scores.

2. Two-Way Univariate ANOVAs. These analyses defined groups based on combinations of categorical variables, usually a demographic variable with treatment or a learning style variable (e.g., gender by treatment or Concept by Education). The dependent variables were stress or performance measures, typically expressed as residualized scores. The univariate analyses were used in the attempts to replicate previous findings pertaining to the AO and as a follow-up procedure for significant multivariate effects that evaluated the effects of demographics.

3. Two-Way Multivariate ANOVAs. MANOVAs were used to explore the possible interactions between treatment (AO vs. Control) and demographic variables (Age, Education, Ethnicity). The typical MANOVA consisted of a two-way classification that defined the between-subjects element of the design (e.g., Treatment x Age) with multiple dependent variables. Typically, the multiple dependent variables were the five residualized stress indicators plus a performance indicator. The performance indicator was either the residualized Academic Test score or the residualized AO test score, depending on the analysis.

4. Test for Parallelism of Regression Lines. The general linear model (GLM) procedure in SPSS-PC was employed in some tests for interactions between intelligence (AFQT) and treatment and demographic variables as predictors of Academic Test performance. These analyses were similar to the two-way ANOVAs, except that AFQT was included in the GLM analysis as a continuous predictor for the dependent variable. The analysis design specified a main effect for AFQT and interactions for AFQT with the classification variables that defined the groups in the analysis. The significance tests for the interactions between AFQT and the grouping variables provides a test of whether the regression lines were close enough to parallel within the subgroups to regard the differences as being the result of chance.

The decision regarding when to use a two-way ANOVA procedure and when to use a two-way MANOVA procedure was determined by whether the analysis was an exploratory test for a relationship. The original project hypotheses asserted that treatment and gender would affect stress and performance. The initial study provided preliminary identification of some additional relationships that were to be replicated. In contrast, the analysis of demographic variables was guided neither by specific hypotheses nor by prior findings. For this reason, it was felt that those analyses should include an omnibus test (Rosenthal & Rosnow, 1984) for the presence of differences before exploring individual indicator variables. The addition of the omnibus test was intended to control for the risk of capitalizing on chance, an outcome that can often occur when large numbers of exploratory tests are performed.

The MANOVAs used Roy's largest root as the omnibus multivariate significance test. This test is one of several reported in the analysis summary. Roy's test was chosen because it was more lenient than the other criteria. A lenient test was desirable because the MANOVA analyses were exploratory. Using stringent criteria might cut off useful lines of investigation prematurely simply because they failed to meet a strict statistical significance criterion. Using a lenient omnibus significance test provided a balance between protecting against experiment-wide error risks that would arise in a series of univariate tests and sensitivity to univariate effects if any were present.

2. Results from First Data Collection Effort, Year Three (Tasks 4 and 5)

a. General Trends for Stress Measures

Table 23 summarizes the basic psychometric information on the stress measures.

Table 23. Descriptive Statistics for Stress and Strain Scales

Scales	No. Items	Pretest			Post Test			r_{tt}	t test
		Mean	SD	α	Mean	SD	α		
Role Clarity	8	3.41	0.54	.732	3.90	0.60	.814	.18	-18.26
Role Conflict	5	3.07	0.53	.480	2.61	0.71	.650	.31	17.44
Negative Affect	6	2.12	0.81	.834	1.90	0.82	.863	.57	8.24
Positive Affect	6	3.11	0.83	.838	3.35	0.88	.850	.57	-8.44
Self-Efficacy	8	3.79	0.52	.576	3.99	0.50	.598	.22	-15.46

Note: Sample size for paired t-tests ranged from 773 to 797, all $p < .001$. "α" is Cronbach's α, an internal consistency estimate of reliability. "r_{tt}" is the test-retest stability coefficient. "t-test" is the paired t-test value for the pre-post comparison.

The major points of interest are as follows:

- Reliability was acceptable. Role Conflict and Self-Efficacy were the exceptions to this generalization. Role Conflict was at least minimally acceptable by research standards ($\alpha \geq .600$) at the post-test time period. Self-Efficacy did not reach this minimal standard at either time. This scale was retained for the analyses despite the low reliability to provide a comparison to the prior study.

- Actual stress at the end of firefighting training was lower than anticipated stress. Role Conflict and Negative Affect decreased. Role Clarity, Positive Affect, and Self-Efficacy increased over time. The effect sizes ranged from ES = .26 to ES = .91. These values are in the small to moderate effect size range.

- Firefighting training affected different people differently. If firefighting training had the same effect on everyone, each person's position in the distribution of scores would remain constant from the anticipated stress to the actual stress. The average level of actual stress could be lower than the anticipated stress, but test-retest correlation would be $r = 1.00$ if there were no measurement errors. The presence of measurement errors reduces the upper limit on the observed correlations to the size of the internal consistency estimates of reliability. Thus, if firefighting training had the same effect on all participants, the test-retest correlations would equal the internal consistency estimates of reliability. This condition was not met for any scale. The maximum test-retest correlation was $r = .57$; the median was $r = .31$. Internal consistency estimates ranged from .48 to .86.

- Results were comparable to those in the initial study (Year Two data collection effort). This statement was true for reliability and changes during firefighting training.

Reliability. In the earlier study, reliabilities in the earlier study ranged from .56 to .82 compared to .48 to .86 in this study. The only notable differences between the studies were somewhat lower reliabilities for Role Clarity (present study, α = .73; initial study, α = .80) and Role Conflict (present study, α = .48; initial study, α = .56). In the initial study, test-retest stabilities ranged from r_{tt} = .21 to r_{tt} = .52 compared to r_{tt} = .31 to r_{tt} = .57 in this study.

Comparison of Anticipated and Actual Stress. In the earlier study, the Role Clarity difference (0.74 points) was the largest difference among the stress indicators. Expressed as an effect size (ES), this change was large (ES = 1.16). Differences for the other stress indicators ranged from .15 to .32 points. The associated effect sizes ranged from ES = .19 to ES = .52. In this study, Role Clarity again showed the largest difference (0.49 points, ES = .91), but the difference for Role Conflict was nearly as large (0.46 points, ES = .87). The remaining three stress indicators showed much smaller differences (0.20 to 0.24 points; ES = .27 to .38). The differences between anticipated and actual stress were highly significant in both studies (initial study, t = 3.01 to t = 18.01, p < .003 for each; present study, t = 8.24 to t = 18.26, p < .001 for each).

b. Gender Differences in Anticipated Stress

The initial study in this project indicated that females and males had very similar perceptions of stress in firefighting training. The findings in this set of data reinforced this initial result (Table 24). The only substantial difference between the two investigations was that the females tended to report higher Positive Affect than males in the earlier study (Female Average = 3.42, Male Average = 3.20, t = -1.95, p < .052), but not in this study (Female Average = 3.10, Male Average = 3.10, t = .00, p < .998). The combined results of the two studies, therefore, indicate that females and males did not demonstrate any reliable differences in anticipated stress at the beginning of firefighting training.

Table 24. Gender Differences in Stress Before Firefighting Training

	Male		Female		t	Sig
	Mean	SD	Mean	SD		
Role Clarity	3.43	.53	3.37	.56	1.38	.169
Role Conflict	3.06	.51	3.04	.55	0.43	.665
Self-Efficacy	3.65	.44	3.68	.45	-0.98	.327
Positive Affect	3.10	.83	3.10	.80	.00	.997
Negative Affect	2.08	.78	2.15	.82	-1.18	.238

c. Stress and Academic Performance

Stress was significantly related to academic performance in this study (Table 25). All of the associations between stress indicators and the residualized academic test scores were statistically significant. All of the associations also exceeded Cohen's (1988) minimum effect sizes criterion (i.e., $r \geq$ .10). In the initial study, only Negative Mood was significantly related to academic performance (r = -.144, z = -2.21, p < .027, two-tailed). That earlier finding was replicated closely in the present analyses (r = -.141, p < .001, two-tailed).

The size of the correlations between stress and academic performance varied across different subgroups of recruits. The associations between stress and performance appeared to differ for males and females. These associations differed between recruits assigned to the control and AO conditions.

The apparent differences in Table 25 are misleading. All but one of the differences were no larger than expected by chance. Given the sample sizes, a female-male difference would have to exceed .145 (absolute) to be significantly different statistically. Only the Self-Efficacy difference exceeded this value (z_{Diff} = 3.01, p < .003, two-tailed). Given the sample sizes for the Control and AO groups, the same criterion was applicable for comparisons between those groups. Only the Role Clarity difference even approached this criterion value (z_{Diff} = 1.93, p < .055, two-tailed).

Table 25. Correlations Between Residualized Academic Test Performance and Residual Gain Scores for Stress Indicators

	Total	Male	Female	Control	AO
Role Clarity	.128**	.160**	.099	.201**	.061
Role Conflict	-.238**	-.161**	-.116*	-.189**	-.093
Self-Efficacy	.171**	.221**	.001	.155**	.092
Positive Affect	.101**	.075	.092	.072	.117*
Negative Affect	-.141**	-.104*	-.151**	-.160**	-.095

*p < .05 **p < .01

Note: Sample Sizes: Males = 391 - 406; Females = 329-338; Control = 317 - 330; AO = 406-413.

Gender Differences in Stress-Performance Associations. The evidence does not make a case for any general effect of gender or treatment on the strength of associations between stress and performance (Table 26).

Table 26. Results of Treatment by Gender ANOVA for Residualized Measures

	Treatment				Gender				Treatment by Gender			
	MS	F	Sig.	Eta^2	MS	F	Sig.	Eta^2	MS	F	Sig.	Eta^2
Role Clarity	5.78	16.38	.000	.020	.19	0.53	.469	.001	.28	0.78	.376	.001
Role Conflict	0.30	0.66	.416	.001	.19	0.41	.524	.001	1.20	2.64	.104	.003
Self Efficacy	1.81	7.89	.005	.010	.22	0.94	.333	.001	.63	2.74	.098	.003
Positive Affect	.68	1.32	.251	.002	4.79	9.34	.002	.012	.21	0.40	.527	.001
Negative Affect	.27	0.59	.443	.001	.25	0.53	.465	.001	.12	0.27	.607	.000
Test Score	0.47	3.20	.074	.004	2.79	18.85	.000	.025	.24	1.62	.204	.002

Note: "MS" is the mean square for the indicated effect, "F" is the value of the F-test, "Sig." is the significance level, and "Eta^2" is the proportion of variance explained. The MS$_{Error}$ terms were Role Clarity

= .35, Role Conflict = .46, Self-Efficacy = .23, Positive Affect = .51, Negative Affect = .46, Abb. Self-Efficacy = .33, and Test Score = .15

In each case, only one of five differences approached or exceeded standard significance criteria. Applying a binomial probability model with $p < .05$ as the probability of obtaining a significant result, at least one of five tests will be significant by chance nearly one quarter of the time ($p = .226$).

The inference that gender and treatment do not affect the strength of associations between stress and performance would have to be modified if any of the differences noted in this study replicated previous findings. The gender difference in Self-Efficacy-Academic Test correlations observed in this study did replicate a similar trend in the previous investigation. In the initial study, Academic Test scores and Self-Efficacy correlated $r = .032$ for males and $r = -.136$ for females. The direction of the difference between the correlations was the same as that in this study. The difference between the correlations was not statistically significant in the initial study ($z_{Diff} = 1.13$, $p < .258$, two-tailed). However, the difference was statistically significant when the results of the two studies were pooled. Computing the pooled probability for the difference across two studies using the method of adding probabilities, the difference would be statistically significant ($\sum p_i = .144$, pooled $p = .010$).

Simply pooling the differences between the Self-Efficacy-Academic Test correlations among males and females is misleading. This approach conceals inconsistencies in the basis for the difference in the two studies. Males produced a significant positive correlation (r = .221) in this study, but the same correlation was nearly zero in the earlier study (r = .032). Females demonstrated a near zero correlation in the present study (r = .001), but had a slight negative correlation in the initial study (r = -.136). The net result is that the difference between males and females is somewhat stable across the two studies, but the results did not indicate a reliable deviation from r = .00 for either females or males. In the absence of a reliable association in at least one of the two groups, it is not logical to talk about a reliable difference between them (Zedeck, 1971). In the final analysis, it was impossible to say that Self-Efficacy was related to performance in one group, but not the other. Until further evidence is available to clarify the size of the correlations within the groups, it would be premature to conclude that a moderator effect is present.

Treatment Differences in Stress-Performance Associations. The absence of reliable differences in the stress-performance associations was even more clear-cut when considering treatment differences. The marginally significant difference between the correlation for the control group (r = .201) compared to the AO group (r = .061) was not evident in the prior study. The corresponding correlations in that study were r = .023 and r = -.037, respectively. Here again, there is no evidence of a correlation that was consistently present in one group, but not the other.

Discussion. The only reliable stress correlate of academic test performance was Negative Affect. Although correlations between stress and performance varied across subgroups in the research design, the variation appeared to be chance. Taking this point into account, the combined evidence from this study and the prior study indicates that Negative Affect is the only reliable correlate of Academic Test performance and that this association applies to all recruits.

It is not clear how the Negative Affect-Performance association should be interpreted. It might be suggested that the worry and depression implied by high scores on Negative Affectivity represent psychological states that interfere with learning. Alternatively, poor performance could be the reason for greater Negative Affectivity at the end of firefighting training. The research design in these studies has been correlational. Recruits have not been randomly assigned to different performance levels or to different mood states. As a result, there is no design basis for resolving the inherent ambiguity of the correlation between Negative Affect and Academic Test performance. The affective state might influence

performance, but the possibility that it is poor performance that produces negative affect cannot be ruled out. In fact, the causal influences could go in both directions at the same time.

d. General Effects of the Advanced Organizer

The first data collection effort in Year Three focused on the Advanced Organizer (AO). The basic analysis was a Treatment (T) by Gender (G) analysis of variance (ANOVA). Treatment groups were the Advanced Organizer (AO) and Control groups. Gender was Female (F) or Male (M). Table 4 summarizes the statistical tests comparing the AO and control groups. The mean scores and standard deviations underlying those statistical tests are given in Table 27.

Table 27. Means and Standard Deviations for Study Design Groups

	AO Females			AO Males			Control Females			Control Males		
	n=	Mean	SD	N=	Mean	SD	n=	Mean	SD	n=	Mean	SD
Role Clarity	212	.11	.49	226	.05	.68	159	-.10	.56	201	-.09	.61
Role Conflict	211	-.07	.65	226	.04	.70	159	.05	.64	201	-.02	.70
Self Efficacy	206	.03	.41	224	.06	.55	159	-.01	.46	200	-.10	.47
Positive Affect	209	-.07	.64	218	.12	.76	152	-.10	.75	195	.03	.71
Negative Affect	209	.01	.68	221	-.05	.65	155	.03	.70	196	.02	.69
Test Score	199	-.06	.40	221	.10	.38	144	-.07	.36	189	.01	.39

The primary observations regarding the effects of the AO were:

- The AO had the same effect on females and males. There were no substantial Gender by Treatment (G x T) interactions. The ε^2 value for this effect ranged from .000 to .003. None of these effects approached even the variance-explained criterion (.01) that Cohen (1988) suggested as the minimum magnitude that would indicate practical or theoretical significance. The small size of the effects was underscored by the fact that even with the present large sample, the largest G x T interaction only approached statistical significance (p = .098).

- Gender differences in stress and strain were limited. Females and males reported comparable levels of stress on four of five indicators (Role Clarity, Role Conflict, Negative Affect, and Self-Efficacy). Women did report less Positive Affect than expected at the end of firefighting training (p = .012; cf. Table 4), but the effect size was modest (ES = .23).

- Female recruits did not perform as well academically as male recruits. The gender difference was statistically significant ($F1,749$ = 18.85, p < .001) and accounted for enough test score variance to fall in Cohen's (1988) small effect size range ($\varepsilon2$ = .025).

- The AO decreased stress and improved performance. Recruits who completed the AO reported greater Role Clarity and Self-Efficacy at the end of firefighting training. These recruits also performed better than expected on Test Performance. These statistically significant AO effects

61

accounted for 0.4% to 2.5% of the variance in the residualized variables. The effect on Academic Test score was the smallest significant effect.

Discussion. The AO improved performance slightly. In contrast to the initial study, the improvement was not evident primarily in females. Instead, the effect applied to the entire recruit population.

The AO effect was small, but the absolute size of the effect may be misleading. Large effects may have been precluded by the study design. The AO treatment was a short preparatory period of instruction preceding a lengthy training process. Recruits who received the AO may have had a head start on those in the control group and subsequently may have learned the firefighting training material more quickly. However, the length of the training may have been sufficient to permit recruits in the control group to catch up to the AO group.

e. Path Model of Stress and Performance

The mechanisms by which the AO influences performance are of substantial interest. Conceptually, the AO was designed to provide a cognitive framework that makes it easier to learn and retain material presented in subsequent firefighting instruction. The AO also may reduce stress by making recruits feel better prepared to master the technical elements of firefighting training and by actually fostering faster learning.

The value of stress reduction as a mechanism for performance was a central concern in the development of this research project. The results reported above set the stage for evaluating the role of stress reduction as a mediator of AO effects on performance.

Rationale for the Path Model. The pattern of relationships between the AO, stress, and performance made stress reduction a plausible mechanism underlying the AO effects. For stress reduction to be a link between exposure to the AO and improved performance, exposure to the AO must reduce stress. If this condition is met and if lower stress is associated with better performance, the reduction in stress produced by the AO would cause improved performance. Stress should predict performance. Further, the stress-performance association should be stronger than the AO-performance association. This prediction follows from the fact that the AO does not completely determine stress reduction, while stress reduction cannot be expected to be the sole determinant of performance. The imperfect transmission of effects at each step in the sequence leads to the prediction that weaker effects will be noted in a two-step process than in a direct influence process.

The general pattern of associations described above conformed to a model with stress as a mediator. AO had stronger effects on stress than on performance. In addition, some stresses were related to academic test performance. This pattern would be expected if the AO influenced stress directly, and stress then influenced test performance.

Test of the Stress-to-Performance Path Model. Detailed analysis of the findings did not support the view that AO effects were mediated by stress. If this were the case, partial correlations between the AO and academic test performance controlling for stress would be zero (Glymour et al., 1987). This prediction was tested with gender partialled out of the associations. The addition of this control variable produced an analysis that paralleled the ANOVA procedures which also held gender constant. With this constraint, the following was true:

1. The AO was a weak, but statistically significant, predictor of residualized Academic Test scores ($r = .066, p < .039$, one-tailed).

2. Controlling for Negative Affect did not account for much of the relationship between the AO and Academic Test scores. The partial correlation (*pr*) was slightly smaller than the bivariate association (*pr* = .062). The partial correlation was statistically significant ($p < .047$), so the residual association still was statistically greater than zero. Although weak in absolute size, the residual association accounted for 88% of the variance explained by the zero-order correlation.

Discussion. Stress did not mediate the association between the AO and improved performance. The statistical tests underlying this statement only controlled for Negative Affect, but the conclusion is valid. The analysis could have been extended to include other stress indicators, but it was not. The reason was that no other stress indicator was a reliable correlate of Academic Test performance when considering both this study and the initial study. Based on this evidence, treating any other stress as a link between the AO and performance in the analysis of data from the present study would produce results that capitalized on chance elements of the present data set.

The conclusion that the AO effect on performance was not mediated by stress reduction increases the plausibility of alternative explanations of the association. Ruling out the stress mediator model increases the likelihood for competing models. Those models include the possibility that the AO works through cognitive mechanisms. However, even if stress reduction is ruled out as an explanation, the results do not prove that cognitive mechanisms are the basis for the AO effect. Other competing explanations can be identified that would also have to be ruled out. For example, the exposure to the AO may simply increase the total instructional time. Also, it is difficult to reach meaningful conclusions by ruling out alternatives. Some direct evidence that the AO provides cognitive organization that facilitates learning ultimately will be needed to verify this mechanism for AO effects. In that light, the present results are most useful in helping rule out stress reduction as a competing explanation.

f. Intelligence and AO Effects

The preceding analyses could be misleading if AO effects depend on intelligence. For example, the ability to develop the mental overview and organizational framework for learning materials embodied in the AO might be part of what makes a person intelligent. If so, the AO would be of limited value to intelligent recruits who already possess the mental tools embedded in the AO. For other recruits, the AO might function as a substitute for the hypothesized organizing effects of intelligence. Two analyses were conducted to evaluate this issue by testing for interactions between AFQT and AO exposure.

Parallelism of Regression Lines. A MANOVA was performed with AFQT treated as a continuous between-subjects factor. Gender (Female vs. Male) and treatment (AO vs. Control) were dichotomous between-subjects factors. Academic Test score was the dependent variable.

The SPSS-PC MANOVA routine was used to test for all effects up to and including the three-way interaction. The interactions involving AFQT assessed the parallelism of regression lines. A significant interaction would indicate different functional relationships between AFQT scores.

The AO did not affect the relationship between intelligence and test performance. The interaction term was statistically nonsignificant for AFQT with Treatment ($F_{1,745} = 0.94$, $p < .333$). The triple interaction between AO, gender, and AFQT also was statistically nonsignificant ($F_{1,745} = 0.07$, $p < .798$). Taken together, these two results indicated that there was no evidence that the relationship between intelligence and performance was different for one group of subjects than another. This conclusion was true for any combination of gender and treatment.

63

Test for Other Forms of Interaction. The preceding ANOVA approach might be insensitive to some types of interactions. Those analyses tested for differences in the slope of linear relationships. If the AO produced curvilinear relationships between AFQT and academic test performance, the analysis might have missed these effects.

Possible intelligence-by-treatment interactions were examined by dividing the sample into deciles based on AFQT scores. A 2 x 10 Treatment x AFQT Level ANOVA then was performed. The ANOVA results included a nonsignificant interaction effect ($F_{9,730} = 1.32$, $p = .223$, $\varepsilon^2 = .016$). The variance explained exceeded Cohen's (1988) minimum for a small effect size if the analysis had involved a single degree of freedom. However, the result was obtained with nine degrees of freedom. The average variance explained per degree of freedom, therefore, was less than 0.2% of the total variance. This value was well below Cohen's (1988) criterion for a study involving that many degrees of freedom.

Visual Search for Interactions. Visual inspection of the trends in Figure 1 showed the expected increase in performance with increasing AFQT.
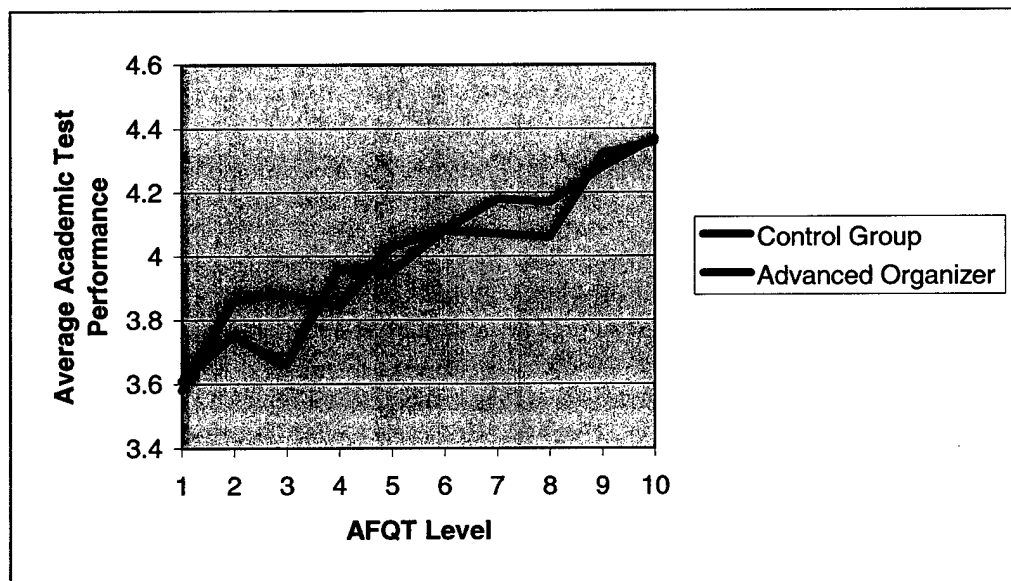
Figure 1. Test Scores for AO and Control Groups by 10 AFQT Levels

The average academic test score for the lowest AFQT group averaged 3.62; scores for the highest AFQT group averaged 4.36. However, differences between the AO and Control groups tended to fluctuate from positive to negative.

The one suggestive generalization in Figure 1 was that the AO might be of more use to recruits of lower than average intelligence. The AO group had somewhat higher scores than the Control group for all three of the lowest deciles. A follow-up test was conducted to determine whether the difference in the less intelligent range was really noteworthy. The AFQT groups in Figure 1 were combined to form low AFQT (1 through 3) and other AFQT (4 through 10) groups. A group-by-treatment analysis then showed a statistically nonsignificant interaction ($F_{1,749} = 1.31$, $p = .252$, $\varepsilon^2 = .002$). Adding gender to the analysis did not affect the results. None of the interactions in the three-way gender x treatment x AFQT level ANOVA were significant ($p > .167$).

Discussion. The AO effect did not depend on the person's intelligence. This statement is the substantive implication of the statistical finding that AFQT did not interact with treatment conditions. The statement generalizes across females and males. While the AO appeared to produce slightly larger effects in some AFQT ranges than others, the variation was within the range expected by chance.

Finding that the AO effect was effectively constant across the range of AFQT scores ruled out a potential explanation for one inconsistency between the present data and previous findings. The initial study in this project indicated that women benefited more from the AO than men; women also had higher AFQT scores than men in that study. Both differences were reversed in this data collection effort. The results of the two studies could have been incorporated into a single general trend if more intelligent individuals benefit more from the AO than less intelligent individuals. However, the straightforward analysis of intelligence as a modifier of AO effects did not support this view.

g. Learning Styles: Interacting with the AO Treatment

The preceding analyses compared the AO and control groups. The remaining analyses focused on comparisons within the AO group. The objective of these comparisons was to determine whether the way a recruit interacted with the AO influenced the effects of exposure to the AO. These patterns of interaction will be referred to here as learning styles to reflect the fact that the behaviors involved reflect both the willingness to be actively engaged in the information acquisition process and the preferred method and content areas of these activities.

Additional Performance Scores. This exploration of learning styles added two new dependent variables to those considered in the overall comparison of the AO and control groups.

AO Test Score. Each recruit who completed the AO was given a knowledge test at the end of the session. The score on this test has been added as one indicator of the effectiveness of the AO. The score was analyzed using both the raw score and the score adjusted for AFQT.

Adjusted Academic Test Score. In the preceding analyses, Academic Test scores were adjusted for AFQT differences. This residualized performance score also is used in the following analyses, but a second adjusted score has been added. The second score adjusts for both AFQT and AO Test scores.

Adjusting Academic Test scores for AO Test performance was important. The initial rationale for developing the AO was to provide a tool that prepared recruits to learn better during subsequent classroom and hands-on training sessions. If the AO performs this function, it is a learning multiplier, something that makes recruits a little better all the time that they are receiving firefighting instruction.

The fact that the AO was designed to be a learning multiplier does not mean that it necessarily is a multiplier. The AO could affect firefighting performance in other ways. For example, the AO could be thought of as just an increase in the total amount of training time. Somewhat better learning would be expected given more time invested in training. Indeed, the effect of the AO might be limited to what it teaches recruits during this initial session. Measuring knowledge status at the end of that session, then statistically removing the effect of that knowledge is a means of evaluating this possible basis for the AO effect.

Performance on the AO Test. AO Test scores were related to both AFQT and Academic Test scores (Table 28).

65

Table 28. Correlations Among the AFQT, AO Test, and Academic Test

Measure	Mean	SD			
AFQT	57.96	17.74	1.000		
AO Test	56.80	15.07	.310	1.000	
Academic Test	4.05	0.44	.495	.346	1.000

Note: N = 413, p < .001 for each correlation.

The results suggest that:

- Real learning occurred while taking the AO. The average AO Test score was 56.8%. This figure is greater than would be expected by chance. How much greater is hard to estimate. Some AO questions had multiple parts, all of which had to be answered correctly to receive credit for the question. For example, several items of clothing might be shown. The recruit then had to indicate which ones should be worn for firefighting.

Other questions used a multiple choice format with four or five options. The general strategy used to estimate the probability of a correct answer by guessing alone was to estimate the probability at 1/k, where k is the number of response options, for the multiple choice questions. The probability of getting a multipart question right was estimated at $1/2^p$ where p was the number of parts to the question. This approach was equivalent to assuming a "yes/no" guess was made for each part.

The basic estimation problem was made more difficult by the fact that choices within some multipart questions were interdependent. Getting one answer right or wrong determined whether other answers would be right or wrong. Analysis showed that, in one case, a question with two parts was answered either with both parts correct or both wrong by better than 99% of the recruits. This item was treated as a single question. A three-part question had low correlations ($r = .22$) between the responses, so each was treated as independent. The probability of getting the item correct by chance was estimated at 0.5^3 (i.e., 0.125). Another question had seven parts. The answers to three parts were perfectly correlated, with the remaining four parts moderately correlated. Treating the item as five independent components produced an estimated probability of getting it right by chance of 0.5^5 (i.e., .03125).

Using the above procedures, a recruit who guessed randomly on every question would be expected to get a score of 31.9% on the test. With this value as the null hypothesis, the observed 56.8% correct translates to $t = 20.07$ (412 df, $p < .001$). The learning represents a proportional reduction in error of 21.9% (i.e., 14.9/68.1).

- Recruits with higher AFQT scores learned more from the AO than those with lower scores. The relationship between AFQT scores and AO Test Scores was modest ($r = .310$), but important if the intent of the AO is to improve performance for recruits who otherwise would struggle. The association could indicate that recruits with high AFQT scores learn more while completing the AO. Alternatively, the association could indicate that more intelligent recruits have better guessing strategies than less intelligent recruits. This strategy could be manifest in both the AFQT test scores and the AO Test scores.

- Learning from the AO is one path to better performance in firefighting training. AO Test scores predicted Academic Test scores ($r = .346$). The relationship still was present after controlling for AFQT. The two predictors combined produced the following regression equation:

$$\text{Academic Test} = 3.068 + (.01073 * \text{AFQT}) + (.00628 * \text{AO Test})$$

The regression coefficients were statistically significant for AFQT ($t = 9.76$, $p < .001$) and for AO Test ($t = 4.85$, $p < .001$). The semipartial correlations (sr) indicated that AFQT uniquely explained 16.6% of the variance ($sr = .408$), while the AO Test uniquely explained 4.1% ($sr = .203$). The total variance explained was 28.6% ($R = .535$). The unique variances explained summed to 20.7%, so the overlap between the AO Test and AFQT accounted for about 7.9% of the Academic Test variance.

- Intelligence and AO learning are additive effects. A moderated multiple regression analysis (Saunders, 1956) was performed to determine whether the slope of the linear relationship between AFQT and Academic Test scores was related to the level of AO Test performance. The interaction term accounted for only 0.1% of the variance in Academic Test scores and clearly was statistically nonsignificant.

h. General Effects of Learning Style

The AO presented recruits with a number of behavioral options. The recruit could, for example, choose between a tutorial approach that presented the information in a linear format or a concept mapping approach that permitted the recruit to explore related branches of the program in an order that he or she chose. Four major components of those interactions could be quantified for evaluation.

- The choice between the linear tutorial and concept mapping had little effect on stress or performance. The tutorial presentation was chosen by most recruits (84.0%; n = 340), but enough chose the concept mapping approach (16.0%; n = 65) to compare the two groups (Table 29). The only noteworthy difference was the difference in Negative Affect. The drop in this stress indicator was much greater than expected in the Concept Mapping group. The other comparisons between the groups were small in absolute magnitude (Difference < .17) and did not approach statistical significance ($p > .390$).

Table 29. Comparison of Tutorial and Concept Groups

Scale	Tutorial		Concept Mapping		Mean	t	Sig
	Mean	SD	Mean	SD			
Role Clarity	.08	.57	.01	.77	.16	.82	.416
Role Conflict	-.02	.69	.00	.75	.03	-.22	.828
Self-Efficacy	.05	.48	.01	.58	.08	.53	.600
Positive Affect	.01	.71	.07	.75	.08	-.62	.538
Negative Affect	.03	.67	-.22	.57	.44	2.79	.006
AO Test	56.27	15.05	57.76	15.22	.10	-.79	.429
Academic Test(1)	.04	.38	.01	.42	.08	.51	.611
Academic Test(2)	.01	.37	-.03	.41	.11	.72	.474

Note: ES = Effect Size. ES is the absolute difference between the mean scores divided by the smaller of the two standard deviations. Academic Test (1) is the Academic Test score adjusted for AFQT; Academic Test (2) is the Academic Test score adjusted for AFQT and AO Test score.

67

- Exploration of gender issues was not critical to the effectiveness of the AO. At several points, recruits could ask questions related to gender issues. The recruits made limited use of these opportunities, but 9.2% (n = 38) of the AO participants made at least one such inquiry. None of the stress, strain, or performance comparisons between those who asked at least one such question and those who asked none was statistically significant (p ≥ .154).

- Recruits who focused on what they needed to remember showed less stress. The AO also provided opportunities to ask for specific guidance regarding what the person had to remember. Only 16.0% (n = 66) of the study participants asked for this information. None of the differences between those who asked one or more such questions and those who asked none approached significance (.137 ≤ p ≤ .179).

- Asking "Why should I care?" had little effect on stress, strain, or performance. Individuals were given the opportunity to ask why certain information was important for them even though damage control was not their primary function. Ten percent (n = 41) of the recruits made use of the opportunity at least once. The stress and performance indicators for those recruits were very similar to those of the recruits who did not make such inquiries (.217 ≤ p ≤ .812).

- Learning style effects were the same for females and males. Thirty-two two-way ANOVAs were performed. Each analysis crossed gender with a learning style variable (i.e., choice, role inquiries, reminders, or gender role) as predictors of one of the eight outcome measures. Only one of the 32 interactions was statistically significant: Role Conflict in the Role Inquiry by Gender analysis (p < .006). One significant result could readily be obtained by chance in 32 significance tests (p = .806 with p < .05 as the significance criterion; p = .175 with p < .006 as the criterion). While these probabilities argue against putting much weight on the finding, role-related inquiries were associated with lower than expected Role Conflict for males (mean = -.23), but higher than expected Role Conflict in females (mean = .24).

- Males chose the concept approach more often than females, but the other interaction patterns were not related to gender. The concept approach was chosen by 21.6% of males (47 of 218), but only 10.2% (20 of 196) of females (χ^2 with Yates' correction for continuity = 8.99, p < .003). The number of males and females was approximately the same for Role Inquiries (20M/21F). Males were more likely to ask what they would be required to remember (35M/27F), but the difference was too small to be statistically significant (p = .516). Males and females were about equally likely to ask gender-related questions (20M/17F; p = .858).

Discussion. This analysis of learning style effects focused on options intentionally built into the AO. These structural details of the AO had little effect on the outcome criteria in this study.

The absence of learning style effects should be interpreted cautiously. One reason is that only a small proportion of recruits used any of the learning style options. Those recruits that did use one or more of the available options were self-selected. The results, therefore, do not indicate what effect the learning options would have if all recruits had been exposed to them. The effects of concept mapping, for example, might be quite different if all recruits were exposed to this instructional approach.

i. Specific Questions in the AO

The database also included markers for 41 points in the AO at which a student could make an inquiry. Inquiries were infrequent overall. However, the relationship between each inquiry and Academic Test performance was examined to see whether any inquiries were relevant to subsequent performance. Because the analysis was exploratory and inquiries were infrequent, a point biserial correlation (r_{pb}) large enough to be p < .10 (two-tailed) was the criterion for identifying an interesting inquiry.

- Five inquiries met the criterion: Q15 ($r_{pb} = -.10$, $p < .045$), Q20 ($r_{pb} = -.09$, p < .062), Q24 ($r_{pb} = -.181$, $p < .001$), Q35 ($r_{pb} = -.084$, $p < .089$), and Q38 ($r_{pb} = -.10$, $p < .051$) as the most informative questions. With one exception, these questions were not among those used to define learning styles in the previous analyses. The specific questions were: "Can you explain the identification system further?" "What happens when an emergency situation is over?" "How does the device work?" "I am not a DC-man. Why is this important for me to know?" "After recruit training, will I be further trained on these systems?". The question "I am not a DC-man. Why is this important for me to know?" was one of the questions used to define the motivational (i.e., "Why should I care?") learning style.

- These associations must have identified individuals with significant performance problems. This inference is justified by the fact that the questions were asked by 3.1%, 1.5%, 2.6%, 1.5%, and 4.4% of the sample for Q14, Q19, Q23, Q34, and Q37, respectively. With this few people to contribute to the differences, explaining even 1% of the total variance was unexpected.

- The different questions tended to identify the same people. Despite the skewed distributions for individual items, ϕ correlations (the product moment correlation between two dichotomous variables) among them were moderate (Range = .08 (Q14 with Q19) to .53 (Q14 with Q23)). Only a single correlation was $\phi < .27$. The average correlation was $\phi = .34$. The sum of the five items produced a scale with moderate internal consistency ($\alpha = .724$).

- Very few people asked even one of these five questions. Ninety-three percent (n = 378) had zero questions. Another 3.5% (n = 14) had only a single question. This left only 3.5% (n = 13) with two or more questions asked.

- People who asked more than one question performed poorly. Because of the skewed distribution of the number of questions asked, the number of questions was used to divide the sample into three groups: 0 questions, 1 question, and 2 or more questions. A one-way analysis of variance comparing the groups was followed by t-tests to contrast the Groups 1 and 2 with Group 0. The results suggested a nonlinear relationship between number of questions and performance (Table 30).

> No Questions. The group with no questions was near the average on all indicators. This result could have been predicted in advance. The sample average represents all the recruits in the sample. Unless there are a few extreme outliers in the data, the average value for a subsample that contains 93.3% of the entire AO treatment group must be close to the average value for 100% of that group.

69

Table 30. Frequency of Questions and Training Outcomes

Scale	Number of Questions				Sig	t_{01}	t_{02}	Sig_{02}
	0	1	2+	F				
Role Clarity	.07	.28	-.03	1.07	.343	-1.32	.59	.559
Role Conflict	-.03	-.11	.29	1.53	.219	.42	-1.71	.089
Self-Efficacy	.05	.07	-.06	.30	.740	-.14	.76	.450
Positive Affect	.03	.01	-.41	2.24	.108	.11	2.15	.032
Negative Affect	-.01	.02	.01	.02	.983	-.17	-.07	.944
Academic Test(1)	.04	.17	-.43	11.68	.001	-1.34	4.57	.000
Academic Test(2)	.01	.15	-.41	9.97	.001	-1.46	4.15	.000

Note: The columns head "t_{01}" and "t_{02}" indicate the t-test results for comparing the 1 and 2+ question groups to the no questions group, respectively. "Sig_{12}" refers to the significance of the difference to t_{02}. Significance levels are not given for the t_{01} differences because none of them were significant at the $p <$ .05 level or beyond. Significance given only if p < .10. Academic Test(1) is the Academic Test score adjusted for AFQT. Academic Test(2) is the Academic Test score adjusted for AFQT and AO Test scores.

> 1 Question. This group tended to fare better than average, but the trends were too weak to be statistically significant when compared to the no questions group.

> 2+ Questions. This group did poorly. The difference was largely limited to test performance and was most evident when adjustments were made for AFQT and AO Test scores.

Discussion. Interrogative interactions with the AO were infrequent. This point merits emphasis. It may not be unreasonable to suppose that people can benefit from this technology only when they make use of the flexible opportunities for learning. If no questions are asked, the AO is another linear presentation of facts. More extensive interaction might have resulted in different effects for the AO compared to the control condition.

The evidence presented here also could indicate that the AO is best used as a simple linear instructional process. Recruits who asked more than one question during the presentation actually did less well than average. On one hand, this observation implies that interactions with the AO identify about one recruit in 30 as a person who will perform poorly relative to his or her mental ability. Such individuals might be targets for additional instruction if resources were available for that purpose. In this case, asking questions could be a self-selection signaling a recruit's awareness that he or she was having trouble in learning the material. On the other hand, it might be argued that the evidence indicates that deviations from linear learning interrupt and impair the learning process. This suggestion should be viewed in light of the fact that the people who chose to ask questions in the present study were self-selected. The apparent effect of that decision may be influenced by who chose to ask. Also, it would be premature to put too much emphasis on the results until they are replicated.

j. Learning Styles and Intelligence

The preceding analyses treated all recruits who chose a particular pattern of interaction with the AO as equivalent. Further consideration suggested that the issue of interactions with intelligence should be revisited at this point. For example, more intelligent individuals might be better suited to using the concept approach to the AO than were less intelligent individuals. Intelligent recruits might have an advantage because they were better able to judge when to ask questions. Alternatively, active engagement

70

with the AO rather than passive observation might be the necessary ingredient to make this tool effective for recruits of below average intelligence. Active interaction might be an indication of motivation to learn, a factor which may be more important for recruits of limited ability than for recruits of high ability. These speculations no doubt represent only some of the possible ways that intelligence might influence the effectiveness of the AO, but they were sufficient to indicate that interactions between style and intelligence were a real possibility.

If intelligence influences the AO effect, MANOVA procedures should show an interaction between AO and intelligence level as determinants of performance. Learning Style x AFQT ANOVAs were performed to determine whether such interactions were present. The sample was split into three groups based on AFQT scores for these analyses. The groups consisted of the lowest 30% of AFQT scores, the middle 40% of AFQT scores, and the highest 30% of AFQT scores. These splits were used rather than the decile splits employed in prior analyses because so few individuals interacted with the AO. Larger numbers in the AFQT groups were needed to ensure that each cell in the analysis contained enough observations to provide stable score estimates.

Performance Effects. MANOVAs to assess performance effects were conducted with the residualized academic test scores and residualized AO test scores as the dependent variables. The interaction was statistically significant for Gender ($F_{2,407} = 7.05, p < .001$) and Role ($F_{2,407} = 3.72, p < .025$) learning style variables. The interaction was nonsignificant for Choice ($F_{2,407} = 0.85, p = .43$) and Remember ($F_{2,407} = 2.13, p = .120$).

The univariate associations were examined to clarify the basis for the significant statistical interactions between AFQT and Gender and Role choices. For Gender, the multivariate effect depended heavily on Academic Test performance ($F_{2,407} = 6.65, p < .001$). The interaction was virtually nonexistent for AO Test score ($F_{2,407} = 0.35, p > .700$). The same was true for Role which produced a significant interaction for Academic Test ($F_{2,407} = 3.70, p < .025$) but not for AO Test ($F_{2,407} = 0.12, p < .892$).

Stress Effects. The fact that learning style might combine with AFQT to influence academic outcomes directed attention to stress reduction as a potential mechanism for the observed effects. A MANOVA was performed with the residualized stress measures and the residualized Academic Test score as the dependent variables. The residualized Academic Test score was included to restrict the analysis to recruits who contributed to the earlier finding of performance differences.

The interaction was significant for Role inquiries ($F_{6,358} = 2.46, p < .025$). However, that result was obtained only because the Academic Test score was included in the analysis. Academic Test score was the only individual variable that showed a significant univariate effect ($F_{2,362} = 3.86, p < .023$). The interaction did not approach significance for any stress indicator ($p > .156$). Similar results were obtained with the Gender inquiries dichotomy, but the stress indicators showed even weaker tendencies ($p > .330$).

Form of the Interaction. The form of the interaction between learning style and intelligence was the same for both style variables (Figures 2 and 3). In each case, low and high AFQT scorers who made one or more inquiries while completing the AO performed below what would be expected based on their AFQT. Recruits in the intermediate AFQT range scored above what would be expected based on their intelligence. The performance of recruits who did not make any inquiries was slightly above average at all three AFQT levels.
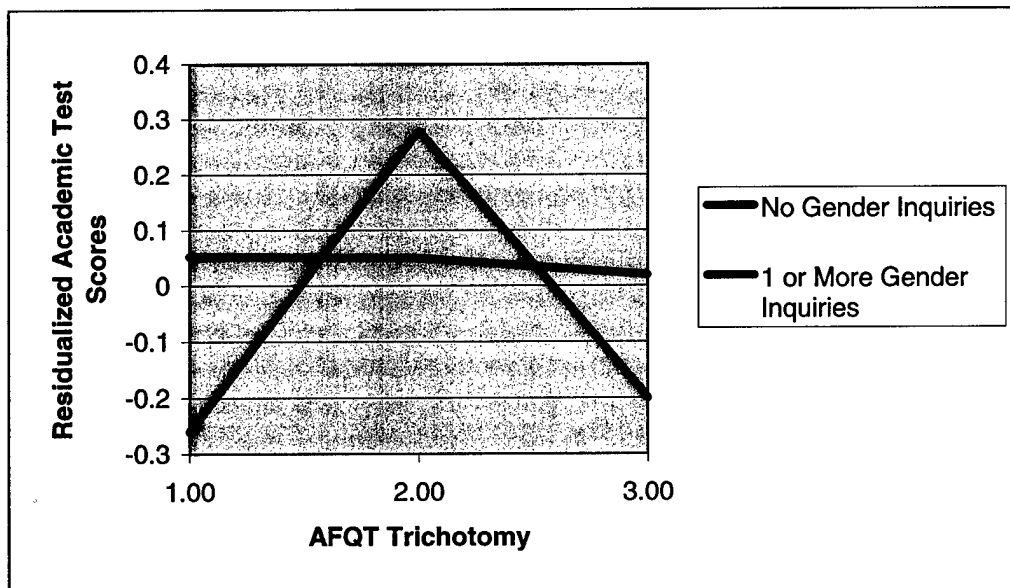
Figure 2. Academic Test Performances as a Function of AFQT and Gender Inquiries
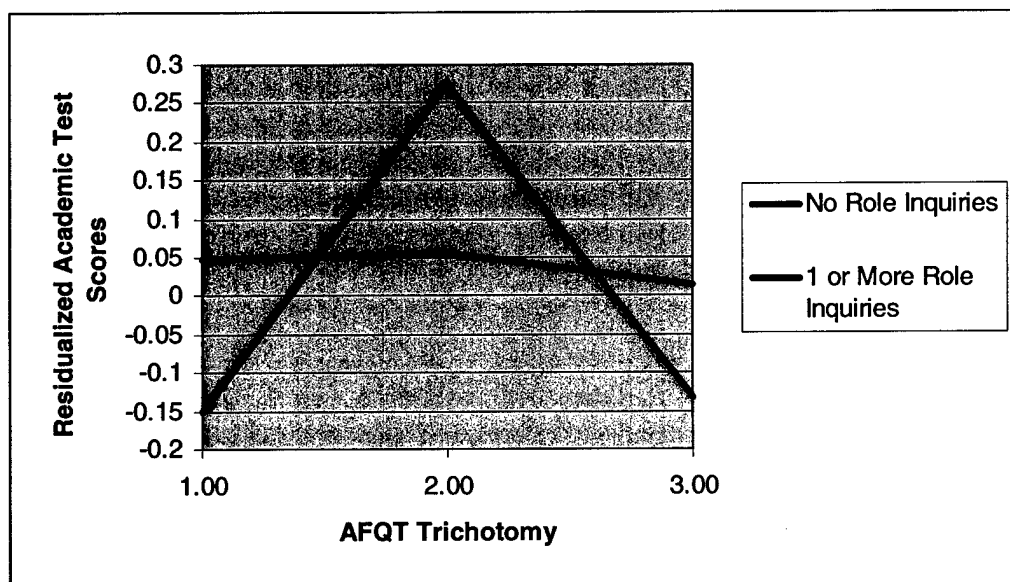
Figure 3. Academic Test Performance as a Function of AFQT and Role Inquiries

Discussion. The significance of a given learning style may depend on the intelligence of the individual being trained. Making inquiries about why it was necessary to learn about firefighting when that was not one's primary shipboard function helped if the recruit was of average intelligence, but hurt if the recruit was in the top or bottom 30% of the AFQT distribution. The same was true for inquiries that focused on gender differences.

The effects of the AO interaction could not be explained by immediate short-term learning. If learning style affected short-term learning, AO Test scores would have been affected. These scores were not

72

influenced. Instead, the inquiries were related only to the longer-term cumulative learning indexed by the Academic Test scores.

The long-term effect of AO interaction was not mediated by stress reduction. If this had been the mechanism for the effect, similar interactions would be expected for the stress response measures. None of the 10 possible interactions for stress responses (2 style variables x 5 stress response indicators) even approached statistical significance.

k. Assessing the Effects of Demographic Attributes

The Year Three research plan included the investigation of demographic attributes as possible influences on AO effects. This portion of this report, therefore, examines age, ethnicity, and education as factors that might affect how a recruit responds to the AO.

(1) Demographic Composition of the Sample

The research design was constructed to provide a sample that was approximately evenly divided between males and females, but no attempt was made to control for other demographic attributes. As a result, the sample variation was rather limited with respect to other demographic attributes. More than half of the study participants fell in a single classification for age and ethnicity, and better than 9 of 10 recruits had 12 years of education (Table 31).

Table 31. Distributions of Demographic Characteristics

	Group	% of Sample
Age	17 years	7.7%
	18 years	53.7%
	19 years	15.9%
	20 years	7.9%
	21-34 years	14.7%
Ethnicity	American Indian	4.0%
	Asian/Pacific Islander	5.4%
	Black	25.1%
	Hispanic	11.8%
	White	52.7%
Education	Less than 12 years	3.8%
	12 years	91.3%
	More than 12 years	4.9%

Some of the groups described in Table 31 were combined in the analyses reported in the remainder of this section. Recruits between the ages of 21 and 34 years old were treated as a single group as in the table. The American Indian and Asian/Pacific Islander groups were combined into a single "Other" group in the analysis of ethnicity. These steps avoided data analysis problems that could have arisen if any of the analysis groups were extremely small.

Small subgroups could have been an analysis problem even though the sample size was large. A group representing 5% of the total sample would have approximately 40 members, but the size of analysis subgroups would be dramatically smaller than 40 in some important situations. For example, the analysis

plans called for testing possible interactions between learning style and demographics. Because only half of the recruits studied went through the AO, the effective sample size was reduced to about 400 recruits. Deviations from the linear presentation of data with no questions asked occurred in only 10% to 15% of the recruits. Thus, only 40 to 60 recruits would choose the less common learning styles. If only 5% of those individuals were Asian/Pacific Islander, for example, that group would consist of 2 or 3 recruits. This number is too small to provide reliable estimates of the effects of the learning style in that demographic category. Even with twice that number in each demographic group, estimates of effects will be imprecise.

The education groups were left as shown in Table 31 even though the two extreme groups were quite small. There was no logical basis for combining the groups to produce larger categories for analysis. The risk of inaccurate estimates of AO or learning style effects, therefore, is greater for the analysis of education effects than for the analysis of other variables.

(2) Associations of Gender with Other Demographic Attributes

Any demographic variable that was correlated with gender could affect the estimation of gender effects on stress and performance. The degree of confounding was examined by cross-tabulating gender with the other demographics. A • ^{2} was computed for each table to determine whether the proportion of females and males was approximately the same in each demographic subgroup. These analyses indicated that:

A. Gender was not related to age (χ^2 = 5.37, 4 df, p > .250).

B. Gender was related to ethnic group membership (χ^2 = 14.49, 3 df, p = .002). Females were the majority of "Other" (52.9%) and Black (55.0%) recruits. In contrast, females were the minority of Hispanic (42.1%) and White (41.6%) recruits.

C. Gender was weakly related to educational status (χ^2 = 5.07, 2 df, p = .079). This weak association arose because recruits with less than 12 years of education were much more likely to be male (73.3%) than were recruits with 12 years of education (52.5%) or more than 12 years of education (55.0%).

These results indicated that gender effects might be confounded with ethnicity effects. There was little risk of confounding gender effects with age or education effects.

(3) Demographic Correlates of AFQT, Stress, and Performance

One-way ANOVAs provided a description of the bivariate associations between the demographic measures and AFQT, stress, and performance. Those ANOVAs indicated that:

- Age was not related to any of the variables (.177 < p < .879).

- Ethnicity was related to AFQT, Role Conflict, and Negative Affect (Table 32). However, except for AFQT, the differences for less than 2% of the variance. Post hoc comparisons indicated that W > O = H > B for AFQT.

- Education was related to AFQT ($F_{2,814}$ = 14.69, p < .01, ε^2 = .034). The major source of differences was the higher average AFQT of recruits with more than 12 years of education (mean = 70.29, S.D. = 17.30) compared to those with 12 years of education (mean = 55.26, S.D. = 17.74) or less (mean = 53.94, S.D. = 14.69).

74

Table 32. Ethnic Group Differences in AFQT, Stress, and Performance

	Whites		Hispanics		Blacks		Other		ε^2
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
AFQT	62.06	17.50	51.34	16.67	45.30	13.64	54.64	16.69	.158**
Clarity	.03	.61	.07	.50	-.06	.65	-.03	.50	.005
Conflict	-.07	.66	-.08	.59	.14	.72	.09	.64	.019**
Efficacy	.02	.49	.07	.37	-.08	.52	.02	.40	.008
Positive	.01	.72	-.17	.79	.06	.70	-.01	.64	.007
Negative	-.06	.64	.12	.66	.05	.74	.08	.68	.010*
Academic	.02	.39	.04	.41	-.05	.38	.01	.38	.005
AO Test	1.41	14.86	-1.33	15.39	-1.78	12.87	-1.82	13.14	.009

Note: "Academic" refers to the score on the academic test adjusted for AFQT. This is an abbreviation of the Academic Test label used elsewhere in this report.

The general pattern of findings was that demographic attributes were not related to stress or performance, but were related to AFQT. In connection with this pattern, it is important to remember that the performance scores were adjusted for AFQT. The fact that performance was not related to demographics does not mean that raw performance did not differ between the groups. The analysis results indicate only that the performance of each group was what would be expected given the measured intelligence in each group.

(4) Demographic Influences on AO Effects

Analyses were conducted to determine whether the AO was more effective for some demographic subgroups than others. MANOVAs were conducted to test for such differential effects. Each MANOVA involved a demographic group x treatment group design. For example, the group classifications in one analysis were defined by dividing recruits based on education and AO condition (e.g., less than 12 years/control, less than 12 years/AO, 12 years/control, etc.). Each of the three demographic variables was paired with treatment condition in a separate analysis.

The analysis was multivariate because six dependent variables were examined in each analysis. The six variables were the five stress indicators and the academic test score. The multivariate test for the demographic x condition interaction was examined first. If this interaction was significant, the univariate interactions then were examined to identify the basis for the significant multivariate interaction. The results were as follows.

Age. The multivariate interaction was statistically significant ($F_{6,688}$ = 4.11; $p < .001$). Significant univariate interactions were found for Role Clarity ($F_{4,690}$ = 3.95, $p < .005$, ε^2 = .022), Role Conflict ($F_{4,690}$ = 2.71, $p < .030$, ε^2 = .015), and Self-Efficacy ($F_{4,690}$ = 2.98, $p < .020$, ε^2 = .017). The interactions for Academic Test performance ($F_{4,690}$ = 2.18, $p < .071$, ε^2 = .012) and Positive Affect ($F_{4,690}$ = 2.08, $p < .083$, ε^2 = .012) approached significance.

The basis for these interaction effects can be seen in the pattern of differences between the AO and control groups for each age group (Table 33). The AO had negative effects on recruits in the youngest age group. Choosing the AO was associated with lower than expected Role Clarity and Self-Efficacy, Positive Affect, and Academic Test performance. Choosing the AO in this age group also was associated with

75

higher than expected Role Conflict. Other groups tended to show the opposite of these effects with the trend particularly pronounced in the 20-year old group.

Table 33. Age and Advanced Organizer Effects

Age	Role Clarity		Role Conflict		Self-Efficacy		Positive Affect		Academic Test	
	C	AO	C	AO	C	AO	C	AO	C	AO
17	.08	-.24	-.06	.12	.11	-.07	.13	-.09	.02	-.07
18	-.05	.15	-.06	.02	-.04	.09	.04	.06	-.02	.05
19	-.19	.11	.16	-.04	-.06	.06	-.28	.07	-.05	.03
20	-.19	.15	.27	-.19	-.31	.08	-.10	-.03	-.15	.10
21+	.04	.07	-.07	-.09	.08	-.04	.00	-.05	.03	-.07

Note: Table entries are group averages for residualized variables. "C" indicates control group; "AO" indicates Advanced Organizer group.

Ethnicity. Ethnicity did not affect the relationship between the AO and stress or performance. The interaction between the treatment conditions and ethnicity was statistically nonsignificant at the multivariate level (Multivariate $F_{3,692}$ = 1.36, p = .229) and for every individual indicator (.273 < p < .778).

Education. Education did not affect the relationship between the AO and stress or performance. The multivariate test for the interaction approached statistical significance ($F_{6,690}$ = 2.09, p < .052). Positive Affect was the only individual measure that approached statistical significance ($F_{2,694}$ = 2.97, p < .052, ε^2 = .009).

Discussion. The effectiveness of the AO did not depend on the recruit's demographic attributes. This conclusion was clearest for ethnicity and education. The interactions for these demographic attributes were statistically nonsignificant in both the multivariate and univariate analyses.

The suggestion that AO effects did not depend on age is based on an evaluation of the overall findings for age. This demographic variable did produce a statistically significant interaction with the AO treatment. The interaction results for age may be discounted on three grounds. First, the significant interactions were more the result of a large sample size than having a strong interaction effect. The interactions explained between 1.7% to 2.2% of the variance in the stress indicators that were significantly affected. This amount of variance explained would represent a small effect size in a single degree of freedom test (Cohen, 1988). However, the age interactions involved four degrees of freedom, so it can be asked reasonably whether the interaction of these effects is large enough to be important. Second, the pattern of effects was complex. The AO was a positive factor for the youngest recruits, had essentially no effect for the modal 18-year-old recruit, then had negative effects for 19- and 20-year-old recruits. A complex model of age effects would be needed to explain this pattern. The complex model would account for only a little variance and would depend on trends that have not been replicated. Finally, the complex model of age effects would not apply to the criterion that was of most interest in this study. Academic Test scores were not involved in the interaction.

The overall conclusion is that there is little reason to believe that the effects of the AO are particularly large or small in any demographic subgroup in the sample. The few statistically significant effects were small and may not replicate. The criterion of most concern for educators, Academic Test performance, did not produce a significant interaction in any of the three analyses.

(5) Gender Differences and Demographics

Gender differences in stress and performance were not widespread in this study, but females did not perform as well as males on the Academic Test. The size of this gender difference might depend on other attributes of the individual. For example, does the male-female test difference increase or decrease as people acquire more education? One might also ask whether the absence of gender differences for stress indicators is the result of differences favoring females in some demographic subgroups combined with differences favoring males in other subgroups.

The impact of demographics on the size of gender differences was not a primary focus of this data collection effort. However, such differences would be an interesting sidelight to the research if they were present. Attempts to improve performance then could focus on why a particular subpopulation performed poorly. MANOVAs comparable to those performed for the AO were conducted with gender replacing the AO as the group classification variable. These analyses indicated that:

Age. The size of gender differences did not vary with age. The MANOVA interaction was statistically nonsignificant (p = .077). The interactions also were nonsignificant (.097 < p < .831) for individual indicators.

Ethnicity. The multivariate interaction was nonsignificant ($F_{6,689}$ = 1.59, p < .149). One univariate interaction was statistically significant (p < .043), but the others clearly were nonsignificant (.394 < p < .790). The single significant univariate interaction was noteworthy because it involved Academic Test scores.

The gender x ethnicity interaction can be described best by comparing the gender differences in performance for the various ethnic groups to the average gender difference in the overall recruit population. On the average, males scored 0.137 points higher on the academic test than females after adjusting for AFQT. The gender x ethnicity interaction arose because the gender difference was much larger than average for the combined American Indian/Asian and Pacific Islander group (.290) and much smaller than average for Blacks (.000). Whites (.153) and Hispanics (.185) had values near the average.

Education. Neither the multivariate interaction test ($F_{6,690}$ = 1.38, p < .223) nor any univariate test (.198 < p < .798) was statistically significant.

Discussion. Gender differences were consistent across different demographic subgroups. Ethnicity may affect the size of the gender differences in academic performance, but this result must be replicated before it can be given much weight. The data in this study did indicate that gender differences in academic test performance were absent among Blacks and were larger than average in the American Indian/Asian/Pacific Islander subgroup. That finding can be placed in proper context by noting that the univariate interaction was 1 of 18 (3 demographic variables x 6 dependent variables) univariate significance tests. The gender-ethnicity interaction for academic test performance was just large enough to achieve statistical significance (p < .043) in the present moderately large sample. The probability of getting one test that was significant at the p < .043 level in 18 significance tests is p = .546. Thus, the observed differences could readily have been the result of chance factors. This interpretation of the results would be refuted if the findings replicate, but replication is needed before giving weight to the observed gender x ethnicity interaction.

(6) Demographics and Learning Style

Two general questions were posed when investigating the relationship between demographic attributes and learning style. First, how do demographic attributes relate to learning style? That is, do different demographic groups have different preferred learning styles? The second general question is "Does learning style affect the rate of learning in different demographic groups?"

Style Preferences. A set of tables was developed showing how many individuals in each demographic group chose each learning style. As in the previous analyses of learning style, style classifications were dichotomous. Four dichotomies were examined: the choice between the linear and tutorial presentation approaches, whether or not the person asked questions about the role of a firefighter, whether or not the person asked questions about what had to be remembered, and whether or not the person made inquiries about gender issues.

The distribution of style choices was examined to determine whether any demographic subgroup was substantially more or less likely than the others to choose a particular learning style. The $\bullet^2$ statistics computed for the demographic x learning style tabulations provided the overall test of whether style choice was associated with demographic status. The results indicated that:

- Age was *not related* to choice of the tutorial approach ($\chi^2 = 1.41$, $p = .842$), role-related questions ($\chi^2 = 1.93$, $p = .749$), reminder questions ($\chi^2 = 1.70$, $p = .790$), or gender questions ($\chi^2 = 1.33$, $p = .856$).

- Ethnicity was *not related* to choice of the tutorial approach ($\chi^2 = 6.54$, $p = .089$), role-related questions ($\chi^2 = 3.53$, $p = .317$), reminder questions ($\chi^2 = 1.42$, $p = .701$), or gender questions ($\chi^2 = 0.32$, $p = .957$).

- Education was *not related* to choice of the tutorial approach ($\chi^2 = 0.87$, $p = .646$), role-related questions ($\chi2 = 1.10$, $p = .576$), reminder questions ($\chi^2 = 1.29$, $p = .524$), or gender questions ($\chi^2 = 1.56$, $p = .459$).

Discussion. Learning style was not related to demographics. Only 1 of 12 χ^2 values even approached statistical significance (p = .088). The likelihood of obtaining one or more results that are significant at the p = .088 level when 12 significance tests are performed is substantial. In fact, a set of 12 significance tests will produce one or more such result this extreme by chance alone about two-thirds of the time (p = .669).

(7) Interactions Between Learning Style and Demographics

Twelve MANOVAs were conducted to explore the possibility that the three demographic variables affected the magnitude of effects for the four learning styles. The dependent variables were the five stress indicators and the academic test score adjusted for AFQT and AO Test score. The adjustment for AO Test score was included to ensure that the comparisons were not biased by the additional learning opportunity provided by completing the AO. Two of 12 interactions between demographic attributes and learning style were statistically significant:

- Choice and Age. The effects of choosing the tutorial approach over a linear style depended on age ($F_{7,355} = 2.70$, $p < .010$). The only significant univariate interaction was that for Self-Efficacy ($F_{4,358} = 3.59$, $p = .007$). The concept map approach had a relatively large positive effect in

recruits who were 21 and older, but had a negative effect on recruits who were 19 or 20 years of age (Table 34).

- Ethnicity and Memory Requirements. The second significant interaction was related to the combination of ethnicity with whether the person asked what had to be remembered. The multivariate effect was statistically significant ($F_{7,356}$ = 3.18, p = .003). Table 35 provides the results describing the weak univariate effect for Self-Efficacy ($F_{3,360}$ = 3.40, p = .018) and a trend for Academic Test score ($F_{3,360}$ = 2.00, p = .114).

Table 34. Self-Efficacy as a Function of Age and Choice of Concept Approach

Age Group	Linear	Concept	Difference[a]
17	-.02	-.18	-.16
18	.06	.20	.14
19	.12	-.20	-.32
20	.15	-.38	-.53
21+	-.13	.27	.40
Total	.05	.07	.02

[a]Computed as Concept - Linear.

Table 35. Self-Efficacy as a Function of Ethnicity and Use of Reminder Questions

Ethnic Group	No	Yes	Difference[a]
Self-Efficacy			
Other	.03	.39	.36
Black	-.07	.38	.45
Hispanic	.14	-.08	-.22
White	.05	.09	.04
Total	.04	.15	
Academic Test			
Other	-.03	-.04	-.01
Black	-.05	.13	.18
Hispanic	.11	-.04	-.16
White	.02	-.11	-.13

[a]Difference computed "Yes" – "No".

Note: Reminder questions were those inquiries that asked which parts of the material had to be remembered.

The Self-Efficacy effect was related to much greater than expected gains in the American Indians/Asians/Pacific Islanders group (.36) and Blacks (.45) coupled with a lower value for Hispanics (-.22). The trend for Academic Test scores was the result of positive effects for Blacks (.18) and Whites (.13) contrasted with a negative effect for Hispanics (-.16).

There was one instance of a statistically significant univariate effect when the multivariate interaction was nonsignificant. This result occurred for Negative Affect ($F_{2,362}$ = 4.01, p = .019) in the analysis of education and the choice between the tutorial and linear approaches to the AO. The multivariate significance test was nonsignificant ($F_{7,357}$ = 1.67, p = .117). The concept approach to the AO was associated with higher than expected negative affect for recruits with less than 12 years of schooling and lower than expected negative affect in all other recruits (Table 36).

Table 36. Negative Affect as a Function of Age and Choice of Concept Approach

Education	Linear	Concept	Difference[a]
< 12 years	-.15	.49	.64
12 years	.04	-.29	-.33
> 12 years	-.11	-.24	-.13
Total	.02	-.22	-.24

[a]Difference was computed as Concept - Linear.

Discussion. Demographics did not interact with learning styles. People from different demographic groups were equally likely to choose any given learning style. Once chosen, each learning style had the same effect for all of the demographic subgroups considered in these analyses.

The second statement is true even though some isolated interactions between demographics and learning style were statistically significant. Those significant interactions should be viewed with caution. First, the multivariate test for interactions was lenient to ensure that even minor effects would be detected. If any other test had been used, none of the multivariate tests would have been statistically significant. Second, only 3 of 84 univariate tests for interactions were statistically significant. With $p < .05$, three or more significant results will often occur by chance alone when 84 tests are performed. Application of the binomial probability model to this case indicates that three or more significant findings would occur in about four out of five cases (p = .797). To make the point another way, the average number of significant results expected by chance would be 4.2, a figure 40% larger than the observed three significant results. Third, one of the three significant univariate interactions implied a rather complex pattern of change in learning style effects with age. The influence of the concept approach to the AO first produced a negative effect for 17-year-old recruits, then a positive effect, shifted back to negative for 19- and 20-year-olds, then back to positive for those over 21 years of age. A complex model probably would be needed to explain this pattern of effects. Obviously, the pattern is not a simple growth or decay model in which the effectiveness of the concept approach increases or decreases with age. If the age-choice interaction is discounted, only one interaction was significant at both the multivariate and univariate levels. This limited effect provides at best a weak basis for claiming that the influence of learning style depends on the learner's demographics.

The conclusion that learning styles have the same impact for people from all demographic groups is subject to one important caveat. Even though the overall sample is moderately large, some of the cells in the interaction analyses were quite small. This situation arose because only half of the study participants were exposed to the AO, and very few of those individuals chose the concept approach or asked particular types of questions. The trends observed here might be important if they were to replicate in other samples.

(8) General Discussion of Demographic Influences

Demographic influences on the effects of the AO were conspicuous in their absence. The overall results of the analyses conducted to determine the nature and size of demographic influences can best be interpreted as indicating that no such influences were present. The isolated instances of statistically significant associations can be discounted as the impact of chance given the large number of significance tests performed. This conclusion applies at the level of both multivariate tests and univariate tests for effects of demographics. Even the few instances of nominally significant multivariate and univariate effects involved small effect sizes. Until there is evidence that one or more of the interactions noted in these analyses is a reliable finding (i.e., until the interaction has been replicated), the best inference from the data is that no important demographic effects were in evidence.

The general conclusion that demographic effects were not present covers three specific conclusions of interest. First, demographics did not affect the magnitude or direction of AO effects. Second, demographics were not related to learning style. Third, demographics did not affect the impact of learning style on stress and performance.

The conclusions from these analyses are contingent on several factors. The learning styles examined almost certainly do not exhaust the range of possibilities for the AO. The demographics that have been considered are commonplace variables, but may not be the critical ones with respect to learning technical material. Perhaps past behaviors (e.g., taking auto shop) or interest indicators would be more appropriate than gender, ethnicity, age, and education. Finally, the choice of criteria may influence the results. Stress and academic performance measured at the end of firefighting training may be insensitive to important effects. As alternatives, one might consider measuring stress levels and knowledge at several points in the training. Perhaps the demographic groups achieve the same outcomes at the end of training, but some subgroups get to that end point faster than others. Another alternative might be the substitution of hands-on performance tests for academic tests. These tests have more immediate real-world implications and may not correspond strongly to academic knowledge. The point to these observations is that only a limited area of the domain of demographics, learning styles, and performance has been directly investigated in these analyses.

3. Conclusions from First Data Collection Effort, Year Three (Task 6)

The first data collection effort during Year Three was undertaken to replicate a previous finding indicating that the AO improved performance in female recruits more than it did the performance of male recruits. This finding did not replicate, but there was evidence that the AO slightly improved the performance of recruits in general. The analyses undertaken to establish this point and to clearly define the AO effects and test for their generalizability across demographic subgroups within the recruit population provided the basis for a number of inferences about the impact of the AO.

- **Academically, women did not perform as well as men in firefighting.** This conclusion refers specifically to the academic portion of the instruction. Academic Test scores were lower for women after adjusting for intelligence. This point is important because the women in this sample had significantly lower AFQT scores than the men. The qualifier "academically" has been added to the conclusion to direct attention to the fact that the women may have performed as well as men in the applied elements of the course.

- **Firefighting training had a minor negative effect on the mental well-being of women.** Women and men did not differ on any stress measure before firefighting training, but women had lower adjusted Positive Affect after training. It must be recognized that this effect is relative. On

the average, Positive Affect increased from the beginning to the end of firefighting training. The increase, however, was smaller on the average for women. The difference might appear to be a result of the relatively poor performance of females, but this interpretation is questionable. Poor performance on the academic test was related to lower stress reduction for all of the stress indicators. If poor performance was the basis for the difference, women would have been expected to score significantly lower than men on all of the stress indicators.

- **The AO slightly improved performance.** The effect was small, but statistically significant. The effect applied to both female and male recruits. Had this not been true, the interaction between treatment and gender would have been statistically significant. It was not. The small size of the effect must be weighed against the fact that the AO is a relatively brief period of instruction, a fact which may limit its impact. The research design only measured the effect of the AO after an extended period of instruction, so the preceding conclusion might be phrased better as "The AO improved performance only slightly over the typical level of learning in the current instructional course."

- **Stress reduction does not link AO exposure to academic performance.** Exposure to the AO was associated with somewhat higher Role Clarity and Self-Efficacy in addition to the previously noted association to better performance. However, if stress reduction mediated the performance improvement, the partial correlation between AO exposure and performance controlling for stress would be much smaller than the simple bivariate association. Only one stress indicator, Negative Affect, has been significantly related to academic performance in both the initial study and the present study. The partial correlation between AO status and academic performance was only trivially smaller than the zero-order correlation and remained statistically significant. This result was not surprising given that the AO did not appear to affect Negative Affect in this study. The two stress indicators that were influenced by AO exposure in this study correlated with academic performance in this study, but they were not related to academic performance in the initial study. Thus, there is no reliable basis for inferring that these aspects of stress can explain the AO effects on academic performance.

- **The AO effect can be explained best by viewing this tool as additional topic-specific instruction.** Scores on the AO Test were high enough to indicate that learning took place as the recruits completed the AO. The learning that occurred predicted subsequent Academic Test scores. The association between AO Test scores and Academic Test scores does not arise because both are manifestations of general intelligence. AFQT predicted both AO Test scores and Academic Test scores, but both associations were moderate in size. AO Test predicted Academic Test performance even controlling for intelligence. This pattern of findings could be interpreted as indicating that general intelligence fosters learning in both the AO and standard instruction, but the specific content of the AO contributes to learning academic material as well.

- **Learning styles were not important influences on stress and academic performance.** Four logically derived learning style variables were investigated. Only one of 24 comparisons between different styles (or, more accurately, the presence of a given style or its absence) was statistically significant. This conclusion would have to be modified if additional research replicated the interaction between AFQT and learning styles observed in this study. Learning styles that involved interrogative interactions with the AO were associated with low academic test performance for recruits at the high and low ends of the AFQT range in the sample. The same learning styles were associated with better than expected performance for the individuals in the middle of the AFQT spectrum. These results were interesting, but should not be given too much weight until they have been replicated.

- **The AO may provide an early warning for identifying recruits who will have problems with firefighting training.** The analyses identified five items in the AO that were associated with poorer academic performance. The five items formed a scale that could identify roughly 3.5% of the recruits as people who would not perform up to their abilities. The performance deficit among those recruits was fairly large, so the AO conceivably could identify a manageable proportion of recruits as at risk for poor performance. The result requires replication before it can be taken too seriously. If the finding does replicate, it would raise the question of whether tools similar to the AO would serve a comparable purpose in other settings.

- **AO effects did not vary across demographic subgroups within the recruit population.** The effects of age, ethnicity, and education were extensively examined. The primary concerns in the analyses were whether these demographic attributes modified the effects of the AO in general or the effects of specific learning styles within the AO. A few isolated statistical interactions were found that might suggest the demographics modified the AO effects. Those interactions were infrequent and could readily have been the results of chance given the large number of significance tests performed. Even if the interactions were not the results of chance, the effect sizes were small. Thus, even if the results were replicated, there is little likelihood of concluding that the AO is particularly potent for any demographic subgroup in the recruit population.

- **Caveats.** The analysis of questions with the AO is subject to some important caveats. First, the groups that were compared were not the result of random assignment. The process of self-selection into the different "conditions" may mean that the questions are only expressions of the person's general behavioral tendencies (e.g., attention problems). Second, the last set of comparisons involving specific questions should be regarded only as exploratory. The criterion that was used to select the questions of interest was also the criterion evaluated to assess the overall effect of the "question index." This circularity was necessary given the nature of the analysis, but caution is appropriate when interpreting the results because of the potential for capitalization on chance. However, the curvilinear nature of the relationship to performance is interesting. One slip can be an accident; repeated problems identify a trend. When the trend is present, performance is poor. That trend is present in only a small proportion of the total sample. This percentage implies that the identification is not important unless these recruits also have substantial other problems. If they do, the small percentage of affected individuals means that instructors can focus on these individuals without an excessive burden.

4. Experimental Methods from Second Data Collection Effort, Year Three (Tasks 4 and 5)

a. Procedures for Implementation of Interventions and Data Collection

Design Efforts Prior to Data Collection

Learning Styles Pretest

Prior to collecting data, a learning styles pretest (see Appendix C) was designed in order to provide specific data regarding which instructional strategies were the most successful at preparing students of different genders, ethnicity, education, intelligence, and age for the academic portion of the course. In the data collection effort during Year Two, students interacted with the AO treatment in a manner that was consistent with their individual learning styles. However, we were unable to discuss the results in terms of statistical significance for two reasons. First, prior to taking the AO, we did not ascertain the recruit's learning style and second, we did not randomly assign recruits to the different instructional strategies

within the AO. AO interaction is defined as selection of method (concept map or linear) and selection of questions within the AO treatment.

Development of CD-ROMs for Each Instructional Strategy

The Advanced Organizer program was reorganized to provide separate programs on CD-ROMs, one for each of the six instructional strategies. The six instructional strategies or "routes" through the AO were:

- Concept Map/No Questions. In this program, the students were in the concept map approach and they were not able to ask the computer any questions.

- Concept Map/All Questions. In this program, the students were in the concept map approach and they were provided with all the questions and answers which were possible in the program.

- Concept Map/Student Selected. In this program, the students were in the concept map approach and they were able to ask questions if they wanted to.

- Linear/No Questions. In this program, the students were in the linear approach and they were not able to ask the computer any questions.

- Linear/All Questions. In this program, the students were in the linear approach and they were provided with all of the questions and answers available in the program.

- Linear/Student Selected. In this program, the students were in the linear approach and they were able to ask the computer any questions.

Preparation of Classroom Instruction

The control group in this data collection effort was provided with one hour of classroom instruction covering basic firefighting content. The instructor at Great Lakes prepared this one-hour block of classroom instruction prior to the data collection effort.

Preparation of Academic Firefighting Test

During this data collection effort, the control and treatment group students were immediately tested upon completion of the AO or the one-hour classroom instruction (test$_1$) and after they completed the firefighting training (test$_2$). A version of the Academic Firefighting test was prepared by the instructors at Great Lakes prior to data collection.

Data Collection Effort

One "integrated" division per day received the AO Treatment. An integrated division consisted of one division of male recruits and one division of female recruits, integrated and processed through basic training as a single division. The procedures for implementation of the interventions and data collection were as follows.

In the morning, the lab was prepared for the first group of recruits. The computers were logged onto and the AO treatments were started and waiting at the sign-on screen for the recruits. When the assigned division for that day arrived, they were brought into a classroom where the consent form was administered.

Recruit participants were introduced to the administrators and topic of study. The consent forms were distributed and read to the recruits. The administrators answered any questions from the recruits at this time. The consent forms were then signed by the participating recruits and collected. The administrators verified that all needed information was on the consent form. One administrator stayed in the classroom to administer tests and one administrator stayed in the lab to administer the AO.

In order to accommodate the number of computers available in the lab and the recruits, groups of 40 recruits at a time were taken from the classroom to the lab to receive the AO. Any recruits not included in the set of 40 would remain in the classroom.

Upon entering the FAST Lab, the recruits were given specific instructions pertaining directly to the AO. The recruits were instructed not to use the "quit" button, as all the data would be lost. At this time, the recruits were instructed to sit at any of the workstations in the lab that had the login screen for the intervention. There were six versions of the AO and they were loaded in equal amounts on the 40 computers in the lab. The recruits sat at the designated computers in a random order and received whichever AO strategy was loaded on that particular computer. The recruits began by taking the 15-item learning styles pretest, which was on the computer, and then they proceeded to take the AO treatment.

The six AO strategies were designed with varying amounts of information given to the recruits. Therefore, the recruits finished at varying times with the average being approximately 55 minutes. When all the recruits were finished, they would return to the classroom and the next set of 40 recruits went to the FAST Lab to receive the treatment.

The recruits that returned to the classroom upon completing the AO took the Firefighting Academic $Test_1$ while the next group of recruits were in the lab receiving the AO. The Firefighting $Test_1$ took approximately 30 additional minutes. This rotation continued until the entire division had received the AO treatment and had taken the Firefighting $Test_1$.

When the last group of recruits had finished the intervention and were taking the Firefighting $Test_1$, one administrator would shut down the interventions, collect the diskettes, log off of the computers, and return to the classroom to help administer the Firefighting $Test_1$, if needed.

When the last group finished the Firefighting $Test_1$, the division in its entirety was released. All data collected would be locked and secured in the lab for the night by the administrators.

The recruits completed the rest of the conventional firefighting course and at the end of that training they received the academic Firefighting Test ($Test_2$) again.

The same procedures were followed for each division receiving the treatment.

The control group was brought to a classroom where the consent forms were administered. The same procedures for the consent forms were followed for the control group as used for the divisions receiving the treatments. Upon completion of the consent forms, the control group received one hour of lecture-type classroom instruction prior to taking the Firefighting Academic $Test_1$. The control group consisted of 25 recruits and could easily be handled in one rotation. Upon completion of the Firefighting $Test_1$, the control group was released. They completed the firefighting course and they took the academic test again ($Test_2$).

b. Analysis Procedures

The analysis procedures consisted of several steps. First, the relationships between potential determinants of academic performance were examined. The objective was to determine whether there was significant risk of confounding the effects of one determinant (e.g., gender) with that of another determinant (e.g., field dependence). This problem would arise if two determinants were highly correlated and would have to be taken into account to ensure accurate attribution of effects in the analyses. The random assignment of people to research conditions was expected to result in weak or nonexistent relationships between group assignment and other attributes, but the efficacy of the assignment process needed verification.

The second step in the analysis created several residualized performance variables. Raw scores on performance tests were adjusted to take into account differences in intelligence as measured by AFQT. Residualized scores adjusting for intelligence were created for the first academic test ($Test_1$) and the final academic test ($Test_2$). The conceptual basis for the AO predicted that this treatment would act as a multiplier for learning during the formal academic part of training. Recruits exposed to the AO would learn more because they would have a better conceptual framework for recognizing and storing factual data. This effect should be independent of any initial learning produced by completing the AO. To isolate this effect of the AO, a second residualized score was computed that controlled for both AFQT and scores on $Test_1$.

The third step in the analysis examined the effects of AO design. This part of the analysis compared the different AO types (linear, concept mapping) and the use of questions (none, student-selected, all) in what was basically a 2 x 3 analysis of variance (ANOVA). The key issue in this part of the analysis was the determination of whether particular elements of AO design affected performance. If some groups were comparable, comparisons between the AO and controls would be simplified. For example, if AO type affected performance, but question utilization did not, and if the type and questions did not interact, all recruits exposed to a particular AO type can be treated as a single group. The resulting focus on comparing just AO types would direct the analysis to the area where differences actually occur, would minimize the risk of capitalizing on chance by reducing the number of significance tests performed, and would increase the power of the analyses by combining recruits into larger groups where appropriate.

The basic ANOVA in this third step was extended to include gender (male/female) and field dependence (independent/dependent) as group classification variables. These extensions made it possible to determine whether the effect of AO design depended on the recruit's gender and field-dependence status. The final ANOVA design, therefore, was a 2 x 2 x 2 x 3 group classification based on gender, field dependence, AO type, and question utilization.

The fourth step in the analysis compared the AO to the control group. The AO groups or levels for this comparison were to be determined by the results of the preceding step in the analysis. Gender and field dependence were retained as variables to assess to determine whether the general effect of the AO (in contrast to specific effects of particular types of AO) depended on these variables.

86

5. Results from Second Data Collection Effort, Year Three (Task 5)

a. Performance Antecedents

This second data collection effort in Year Three examined four variables that might explain differences in firefighting performance: gender, AFQT, learning style, and experimental treatment. For the purposes of this analysis, experimental treatment was an eight-group classification consisting of the controls group, the six AO designs, and a set of individuals with indeterminate AO treatment status. The last group was necessary because computer malfunctions resulted in lost data for a substantial proportion of the recruits studied (n = 67).

1. Males and females were evenly distributed across treatment groups ($\chi^2 = 2.06$, 7 df, p = .956).

2. AFQT scores were comparable across the treatment groups ($F_{7,271} = 0.49$, p = .841).

3. Learning style was mildly confounded with treatment group ($\chi^2 = 10.05$, 5 df, $p = .074$), largely because of a higher rate than expected of field-dependent individuals in the Linear-All Questions AO condition (65.5% vs 40.9% in the other conditions).

4. The average male had a slightly higher AFQT score (62.1) than the average females (58.5), but the difference was statistically trivial (point biserial $r = -.09$, $t_{277} = 1.58$, $p = .116$).

5. Females were more likely to be field-dependent than males (60.5% vs 51.4%), but the trend was not significant ($\phi = .09$, $\chi^2 = 1.50$, 1 df, p = .221).

6. Learning style was weakly, but significantly, related to AFQT (Dependent = 63.9, SD = 19.9; Independent = 57.7, SD = 19.6; point biserial r = -.15, $t_{278} = 2.08$, p = .039).

Overall, there was no substantial confounding among the potential performance antecedents.

b. Performance Measures

The first ($Test_1$) and second ($Test_2$) academic tests of firefighting knowledge were the performance measures in this study. Previous data collection efforts in this project included performance on a test that was included in the AO as a performance assessment. Computer problems produced substantial missing data for this measure in the present sample, so AO test scores were not included in the primary performance analyses. However, these measures were included in measures reported in Table 37. AFQT, gender, and learning style were included in the table to show the simple bivariate associations between these variables and performance.

Table 37. Associations Among Performance Measures and With AFQT

AFQT	1.000				
AO Test	.407**	1.000			
FF Pre-Test	.387**	.425**	1.000		
FF Test	.498**	.429**	.356**	1.000	
Gender[a]	-.094	-.299**	-.215	-.121*	1.000
Field Dependence[b]	-.154	-.058	-.097	-.090	.090

Gender: Male = 1, Female = 2
Field Dependence: 1 = Dependent, 2 = Independent

Note: N = 240-279, except for AO test, N = 152-161. *p < .05 **p < .01

Performance measures were moderately correlated to one another (.356 ≤ r ≤ .429), but were slightly more strongly related to AFQT (.387 ≤ r ≤ .498). Gender was a weaker predictor of performance, but still consistently produced associations that were large enough to be statistically significant and fall in Cohen's (1988) range of small effect sizes. Learning style was not significantly related to performance.

c. Effects of Missing Treatment Data

Recruits with missing treatment data were compared to those with treatment data to assess the possibility that the missing data biased the study findings. Because all of these individuals had originally been assigned to one of the AO conditions, the effects of the missing data were assessed by comparing this group to those conditions. The analyses indicated:

1. Recruits with missing treatment data did not differ from the average recruit with known AO status (p > .531).

2. Recruits with missing AO data did not perform differently than recruits in any of the specific AO conditions (.098 < p < .941).

These findings suggested that the missing data did not distort the effects for any of the treatments.

d. Residualized Performance Scores

The relationships between AFQT and the scores on the academic firefighting tests were anticipated on the basis of previous research in this project and evidence from other research indicating that general intelligence predicts academic performance (Ree & Earles, 1991). Adjusted academic performance scores, therefore, were computed to remove the effects of differences in general intelligence from subsequent comparisons of groups within the study. Residualized scores were computed based on the following regression equations:

$$Test_1 = 1.952 + (.0123*AFQT)$$
$$(Mse = .321, R = .387, R2 = .150, R2adj = .146)$$

$$Test_2 = 3.385 + (.0107*AFQT)$$
$$(Mse = .127, R = .498, R^2 = .248, R^2adj = .245)$$

The residualized score for each equation was the difference between the predicted value and the observed score on the dependent variable.

Additional residualized scores were computed to allow for the effects of gender on $test_1$ performance and early learning (i.e., $Test_1$) on final academic test performance ($Test_2$). These residualized scores were computed by regressing the relevant dependent variables on AFQT and the additional predictor, then taking the difference between the observed performance and the predicted performance. The regression equations were:

$$Test_1 = 2.296 + (.012*AFQT) + (-.218*Gender)$$
$$(Mse = .311, R = .424, R^2 = .180, R^2adj = .174)$$

$$Test_2 = 3.132 + (.128*Test_1) + (.009*AFQT)$$
$$(Mse = .124; R = .534, R^2 = .286, R^2adj = .280)$$

Note that gender was added to the $Test_1$ equation based on the ANOVA findings described in the following section of this report. Gender was treated as a simple dichotomous predictor in this equation with females scored as '1' and males as '2'.

e. Analysis of AO Effects on $Test_1$ Performance

The next analysis focused on AO design effects. A four-way 2 x 2 x 2 x 3 ANOVA design estimated the effects of gender (Male, Female), learning style (Dependent, Independent), AO type (Linear vs. Concept Map), and question utilization (All, None, or Choice). The analysis employed the first test score adjusted for AFQT ($Test_1$) as the dependent variable. The results, as shown in Figure 4, indicated that:

1. Gender produced the only significant main effect on adjusted performance ($F_{1,148} = 4.11$, $p = .045$). Men scored higher than women (Male average = .23; Female average = -.19). No other main effect approached significance ($p > .349$ for each).
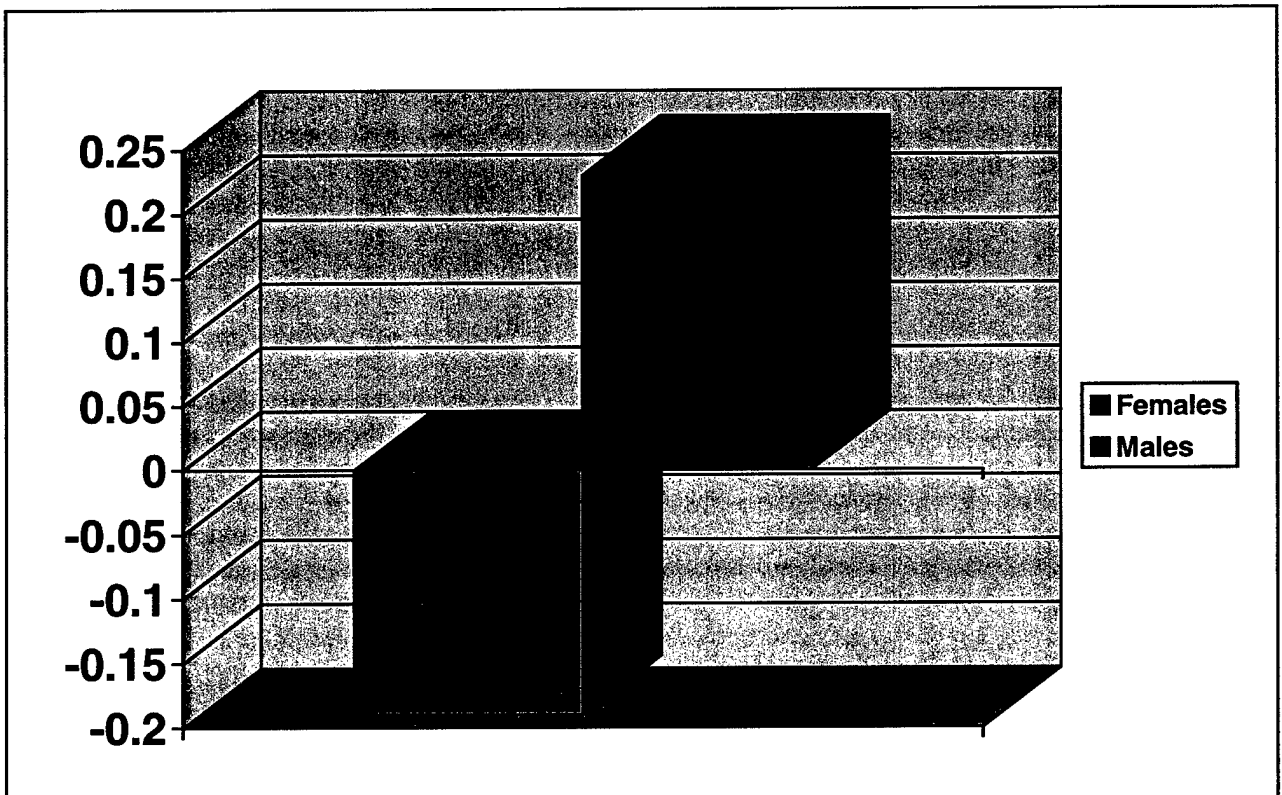
Figure 4. Females and Male $Test_1$ Performance Adjusted for AFQT

Note: Data are standardized residuals for pretest performance regressed on AFQT, n=74 for males and n=107 for females.

89

2. Two interactions were statistically significant: AO type x Gender ($F_{2,148}$ = 5.58, p = .005) and Question Utilization x Learning Style ($F_{2,148}$ = 3.41, p = .036). All other interactions were statistically nonsignificant (.058 < p < .850).

The ANOVA model of the $Test_1$ performance indicated by this analysis, therefore, consisted of a main effect of gender combined with interactions between the use of questions and both the type of AO and learning style status.

f. AO Type and Question Utilization

The significant interaction between AO type and question utilization was evaluated further by simple effects analysis (Winer, 1962). The $Test_2$ measure that adjusted for gender was used in these analyses because the interaction was identified in an analysis that controlled for gender. Figure 5 illustrates these results.
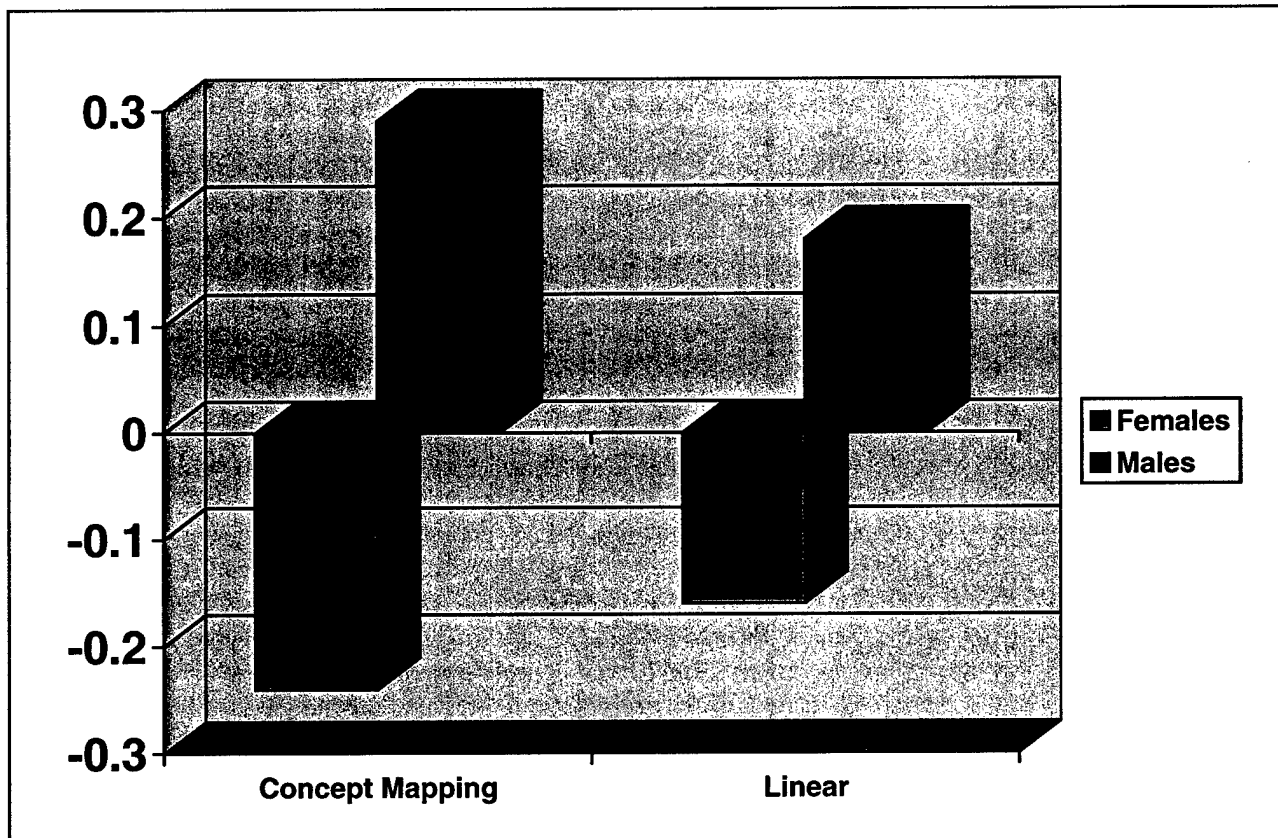
Figure 5. Gender and AO Type Interaction for $Test_1$ Performance

Note: Dependent variable is the standardized academic test score adjusted for AFQT and Pretest score. Group sizes were C=22, L/N=31, L/A=28, L/S=28, C/N=28, C/A=24, and C/S=32.

Question utilization affected performance for the Concept Map AO type ($F_{1,85}$ = 3.15, $_p$ = .048, ε^2 = .069), but not the Linear AO type ($F_{1,93}$ = 0.87, p = .422, ε^2 = .019). Planned comparisons were evaluated to dissect the Concept Map differences. These comparisons contrasted the "No Questions" condition with the other two conditions. Letting the students decide which questions resulted in slightly poorer

performance relative to the no questions reference point ($d = -.24$, $p = .083$). Forcing the recruit to see all questions produced essentially the same performance as obtained with no questions ($d = .10$, $p = .501$).

The interaction also was examined by holding the question status of the participants constant and examining differences between the AO types. AO types did not differ when no questions were asked ($t = -1.15$, 58 df, $p = .257$) or when all questions were asked ($t = -1.14$, 55 df, $p = .261$). The difference approached significance when students decided which questions to ask ($t = 1.99$, 62 df, $p = .051$). In this latter case, students in the Linear AO condition did better than students in the Concept Map condition (L, 0.15, SE = .16; CM, -.30, SE = 0.15).

Taking all of the comparisons into consideration, the least structured AO produced poorer than average performance. The designation "least structured" is suggested by the fact that both the order of presentation of the information and the questions to be asked were left up to the recruit.

g. *Learning Style and Question Utilization*

Question utilization affected performance among field-dependent individuals ($F_{2,74} = 3.11$, $p = .050$), but not among field-independent individuals ($F_{2,92} = 1.47$, $p = .235$). Among field-dependent recruits, the performance of recruits in the "no questions" treatment was virtually identical to that of the recruits in the student-selected questions treatment ($d = .013$, $p = .940$). However, field-dependent recruits who were forced to examine all the questions and answers performed better than those in the "no questions" treatment ($d = .33$, $p = .035$). Although the overall comparison between the groups was nonsignificant (Figure 6), it was of interest to note that forcing field-independent recruits to examine all of the questions had the opposite effect ($d = -.27$, $p = .090$).

91
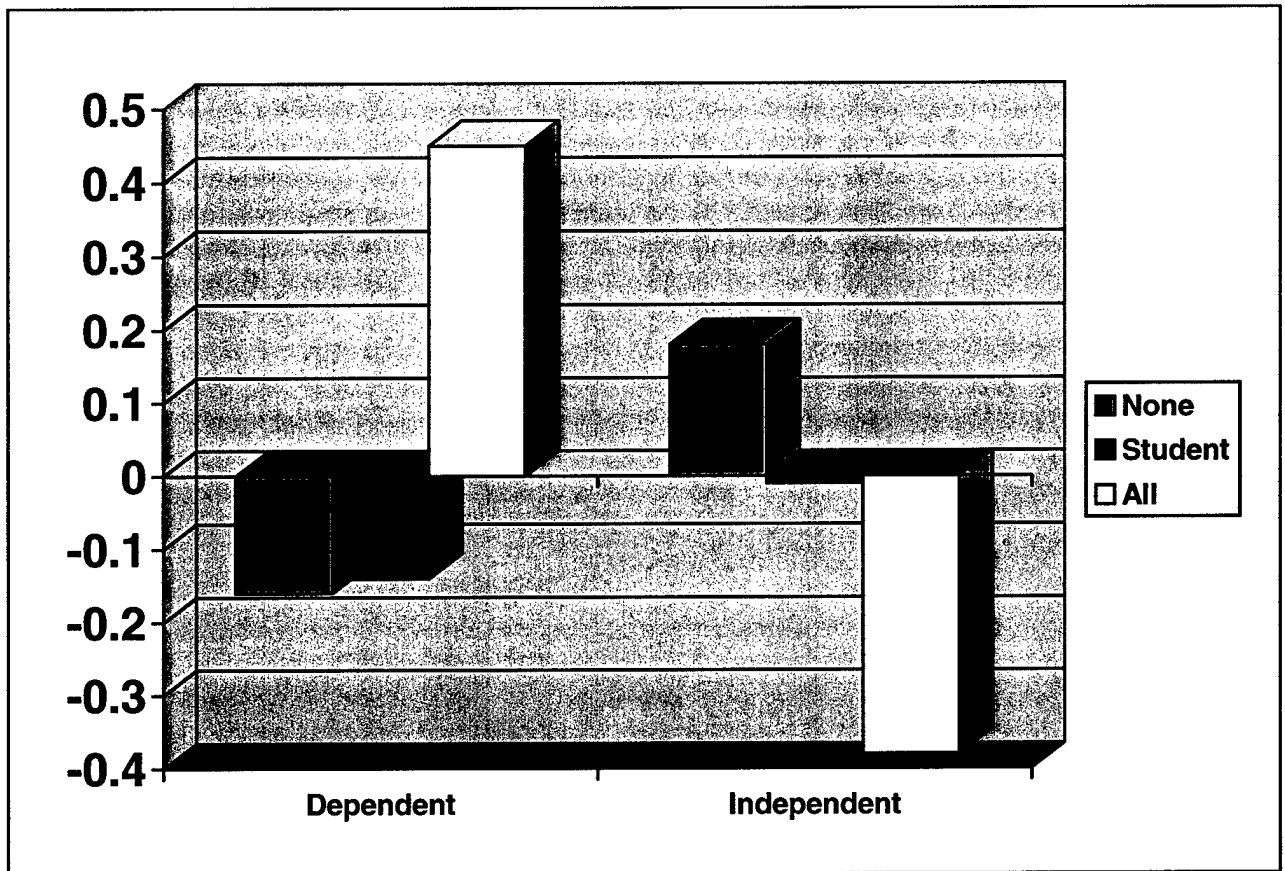
Figure 6. Learning Style x Question Utilization Interaction for Pretest Performance Adjusted for AFQT

Note: Data are standardized residuals for pretest performance regressed on AFQT. Cell sizes were Dependent/None = 23, Dependent/Student = 23, Dependent/All = 31, Independent/None = 36, Independent/Student = 39, and Independent/All = 20.

The same trend was evident when the question utilization was held constant while the effects of learning style were examined. The comparison for the recruits exposed to all questions indicated significant performance differences between field-dependent and field-independent recruits (t = 2.58, 49 df, p = .013). The comparison between the two learning style groups for recruits in the student-selected questions treatment (t = -0.72, 60 df, p = .476) and the no questions treatment (t = -1.20, 57 df, p = .236) did not approach significance.

The overall pattern of differences was clear. Forcing the student to listen to all the questions and their answers improved performance for field-dependent recruits, but tended to impair performance for field-independent recruits.

h. Effect of AO on Test₁ Performance-Contrast with Controls

The presence of two interactions in the preceding analysis raised doubts about the propriety of making contrasts between aggregated groups (e.g., "all", "no" questions recruits) and the control group. Therefore, the comparisons to controls was conducted as a one-way ANOVA with planned contrasts. The

92

group classification consisted of the control group and the six AO designs. The planned contrasts compared each AO design to the control group, as shown in Figure 7.
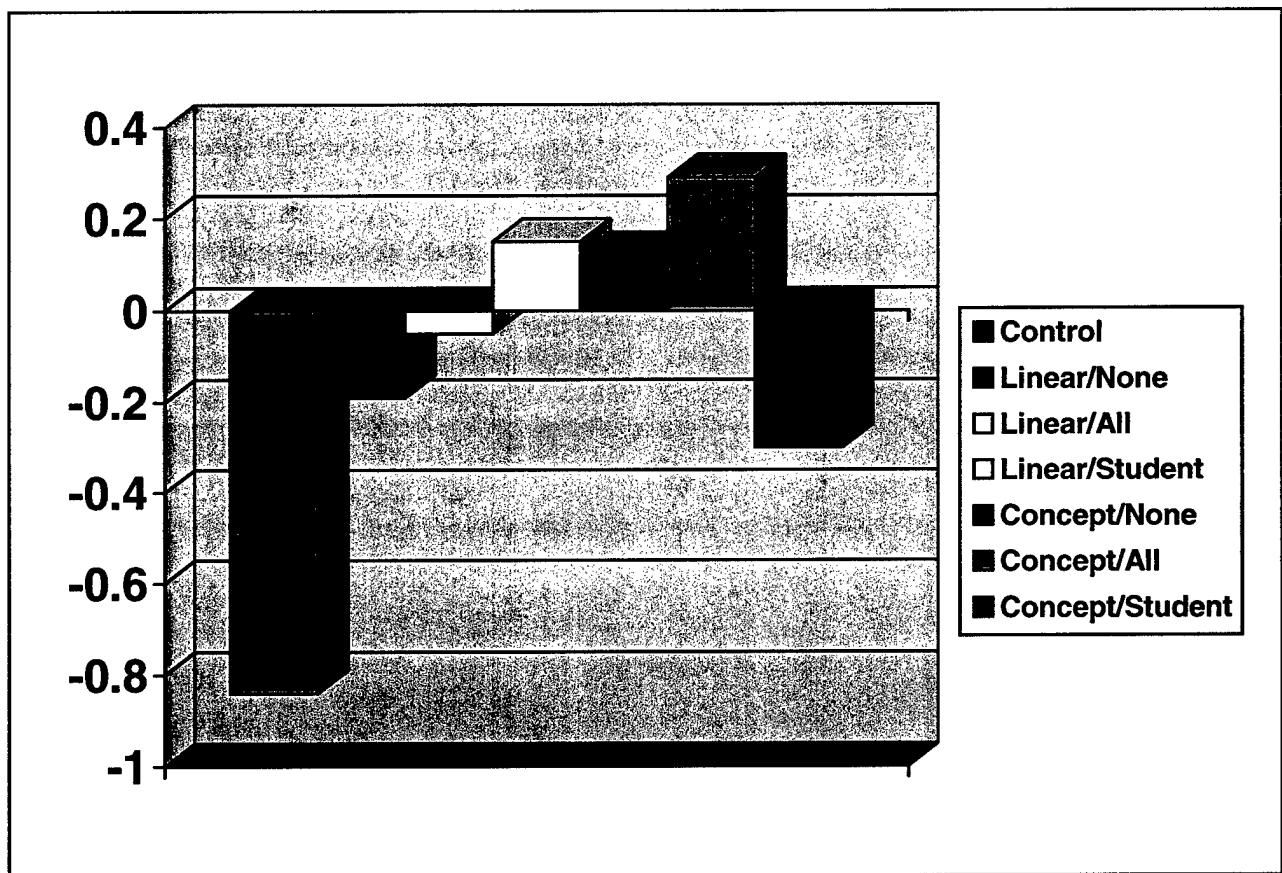
Figure 7. Comparison of AO Designs to Control Group for Test₁ Adjusted for AFQT and Gender

Note: Dependent variable is the standardized pretest score adjusted for AFQT and Gender. Group sizes were C=23, L/N=31, L/A=31, L/S=31, C/N=29, C/A=26, and C/S=33.

The Test₁ score for each AO condition was higher than that for the control group. Contrasts ranged from .329 to .652; every pairwise comparison was statistically significant ($.001 < p < .033$). The results were very similar when Test₂ was analyzed to assess the effect of controlling for gender. Contrasts ranged from .311 to .650; significance tests ranged from $p = .039$ to $p < .001$. The differences between conditions are illustrated for Test₂ in Figure 7.

Based on the initial AO-control contrasts, the AO groups were combined into a single category. A two-way ANOVA was conducted with AO vs. Control as one classification variable and gender as the other classification variable. The interaction terms in these analyses were examined to see whether the size of the AO advantage over the controls was related to gender. There was virtually no difference in the size of the effect for women and men ($F_{1,203} = 0.07, p = .797$).

93

Summary of Test$_1$ Performance

The analysis of firefighting Test$_1$ scores produced three major findings. First, every AO design was superior to the control group. Second, the AO did not eliminate gender differences. Female recruits still showed lower performance than male recruits. The difference was moderate in size by Cohen's (1988) criteria (d = 0.42). Third, AO design had limited performance effects. The only guideline that was suggested by the analysis was the possibility that at least some structure should be imposed on the learning experience. Recruits who were allowed to choose which questions to ask in the concept map approach did relatively poorly on Test$_1$ performance. Finally, the AO effects were not substantially modified by gender or learning style. The influence of forcing recruits to examine all possible questions was positive for field-dependent learners, but weakly negative for field-independent learners.

i. Cumulative Effects in Firefighting Training

AO concepts include the possibility that exposure to an AO will amplify the effects of subsequent instruction. Students exposed to an AO would be expected to learn more efficiently in subsequent instruction because the AO materials provide a framework that permits more effective learning of firefighting material. What might otherwise appear as isolated facts might be given sufficient structure to help the recruit organize and recall facts later. The effect might be akin to the difference between learning nonsense syllables and learning meaningful prose.

Previous results in this series of studies suggested that if the AO has an impact on learning, the cumulative effect was modest. Initially it appeared that the AO helped reduce male-female differences in performance, but later work suggested that this was not the case. The best view at the time the first data collection effort in Year Three was initiated was that the AO had modest effects for both men and women. The present study was designed to examine the hypothesis that the AO effects were produced by early learning, measured by Test$_1$, and to explore the effects of imposing different types of AO structure on the students.

The analyses reported in this section attempted to replicate the prior findings regarding the impact of AO on performance in overall firefighting training. The analyses attempted to extend those prior observations by testing the hypothesis that the final performance differences reflected no more than the effects of initial learning resulting directly from the exposure to the AO. The effects of differences in AO design also are examined.

Two dependent variables were used in the analyses to address the question of how the AO improves performance. One dependent variable (Acad$_1$) is the final academic test score (Test$_2$) adjusted for AFQT differences. The same dependent variable was analyzed in the previous studies in this project. The second dependent variable (Acad$_2$) adjusts the Test$_2$ score for both AFQT and the Test$_1$ score. This dependent variable uses Test$_1$ to estimate the immediate learning effect of exposure to the AO. Adjusting for this immediate learning results in a residualized measure that reflects subsequent learning adjusted for general intelligence.

j. Effects of AO Design

Two four-way analyses of variance were conducted to determine whether there were differences between AO designs. The 2 x 2 x 2 x 3 group classification in these analyses was based on gender, learning style, AO type, and question utilization. Separate analyses were carried out for Acad$_1$ and Acad$_2$ as follows:

94

1. *Academic Test Performance Adjusted for AFQT (Acad₁).* The main effects of AO type and question utilization were nonsignificant. Neither effect accounted for as much as 1% of the variance in adjusted academic test performance ($\varepsilon^2 < .009$). Only one of the remaining effects even approached statistical significance (AO Type x Gender, $F = 2.51$, $p = .115$, $\varepsilon^2 = .017$).

2. *Academic Test Performance Adjusted for AFQT and Test₁ Performance (Acad₂).* The AO type and question utilization main effects were weak ($\varepsilon^2 < .014$) and statistically nonsignificant ($p > .407$). No interaction effect was significant ($p > .125$).

k. Control Group Comparisons

The comparisons between the control group and the AO groups were made at two levels (Figure 8). First, comparisons were made for each of the six AO configurations considered separately. This set of comparisons was included to parallel the earlier analysis of the Test₁ scores. Second, all of the AO groups were combined into a single category for comparison to the controls. This approach was justified based on the immediately preceding analyses that showed no significant effects of AO type, question utilization, gender, or learning style. Combining the AO groups provided a larger sample size that would increase the power of the statistical tests. Each analysis was conducted first with the academic test score adjusted for AFQT (Acad₁) as in prior studies, then with the academic test score adjusted for AFQT and for Test₁ (Acad₂). The analysis results indicated:

1. AO did not improve performance when Acad₁ was the criterion. Pairwise contrasts ranged from $d = -.02$ to $d = .12$ with SE between .099 and .103. The largest d was equivalent to a small effect size (ES = 0.34) when compared to the within-group standard deviation in the analysis.

2. AO design did not improve performance when Acad₂ was the criterion. In fact, the AO groups had poorer performance than the control group in five of six comparisons. None of the differences approached significance ($.450 < p < .967$).
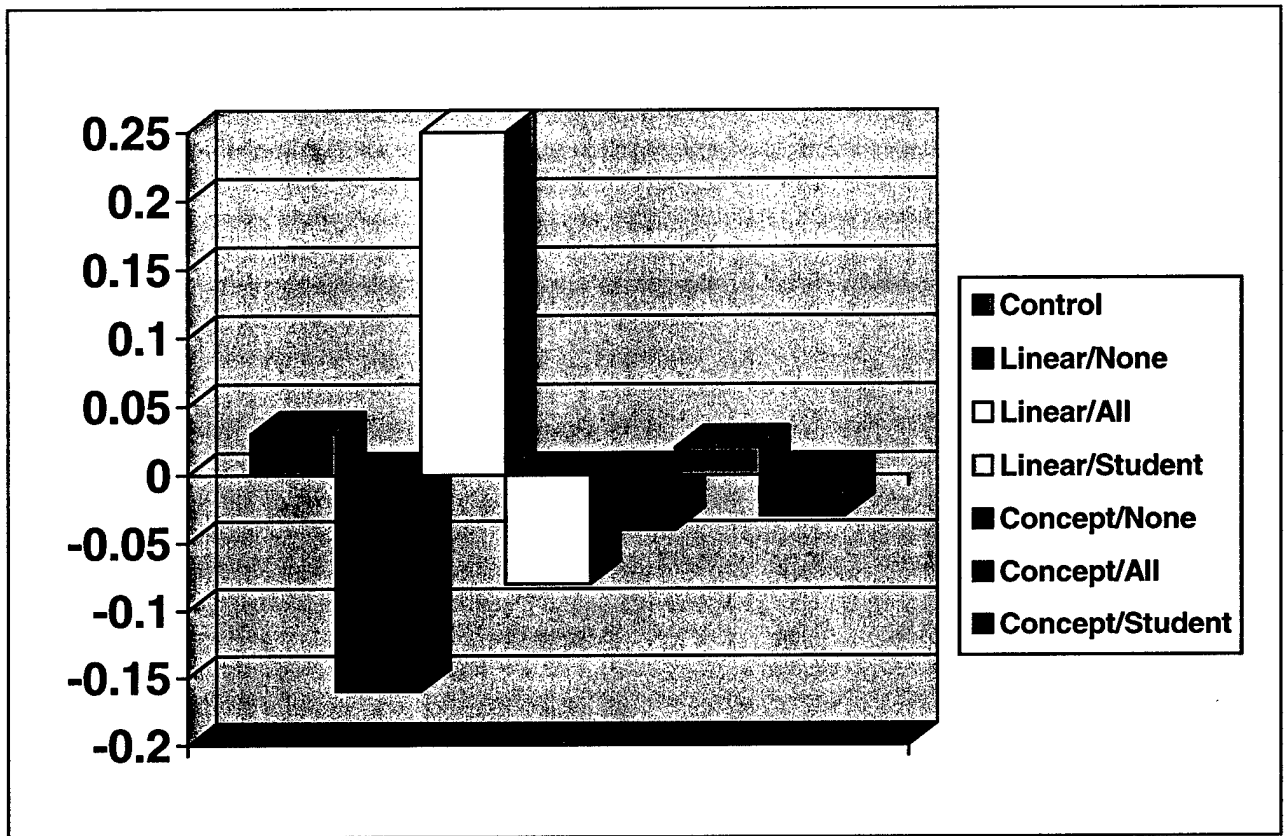
Figure 8. Comparison of AO Designs to Control Group for Test$_2$ Adjusted for AFQT and Test$_1$

Note: Dependent variable is the standardized academic test score adjusted for AFQT and Pretest score. Group sizes were C=22, L/N=31, L/A=28, L/S=28, C/N=28, C/A=24, and C/S=32.

3. When AO designs were combined to form a single group, the difference between the performance of recruits exposed to the AO and the controls remained nonsignificant despite the increased power provided by the larger sample size (Acad$_1$, t_{191} = 0.18, p = .855; Acad$_2$, t_{198} = -.58, p = .564).

4. The size of the AO effect was not affected by gender or learning style. Two-way ANOVAs tested for gender and learning style effects on the contrasts between AO and the control group. If such differences existed and were large enough to be important, the result would be a significant interaction between gender and treatment group in the corresponding ANOVA. The gender interactions were statistically nonsignificant for Acad$_1$ ($F_{6,186}$ = 0.59, p = .742) and Acad$_2$ ($F_{6,179}$ = 0.62, p = .717). The learning style interactions were even weaker for both Acad$_1$ ($F_{5,160}$ = 0.52, p = .759) and Acad$_2$ ($F_{5,153}$ = 0.30, p = .912).

Overall, the AO had no significant effect on final academic test performance (Test$_2$) when compared to the study control group. The absence of significant effects could not be dismissed as a product of small sample sizes. The AO group had lower scores after adjusting for both AFQT and Test$_1$. If this trend held in a larger sample, the inference would be that AO impaired learning in the academic classes that are part of the standard recruit training procedures. The best estimate is that the AO had no effect on final academic performance compared to the study control group.

l. Comparison to Untreated Controls

The conclusion stated in the preceding paragraph was at odds with previous findings in this project. The initial investigation in the project suggested that the AO helped recruits, particularly female recruits. The results of the second study suggested that the AO helped recruits in general. The absence of any effect at all in this study is at odds with those findings, so it is important to consider why the results were not consistent with prior findings.

One reason that the present results may differ from the prior findings is that the control group had a different character in this study. In the previous studies, the control groups were recruits who received standard academic instruction only. The experimental groups received the same instruction plus the opportunity to learn from the experimental treatment. The control group in the present study was different in that these recruits received an extra period of instruction. In effect, the control group was a traditional academic approach group matched to the AO recruits for total instructional time.

The change in the nature of the control group left open the question of whether the AO produced improved performance relative to recruits who merely received the standard educational program. Data for individual recruits in a comparable control group was not available, but aggregate data for the two divisions from which the present control recruits were drawn was available (Figure 9). Those two divisions consisted of 71 males and 54 females. Average scores on academic tests are routinely recorded in basic training and were available for those two divisions. The average male score on the academic test was 4.00; the average female score was 3.72.

The aggregate data was used to estimate the performance of a control group comprised of recruits who experienced only the standard academic training. The total test score cumulating across all recruits in the two Recruit Divisions was computed by multiplying the average score for each Division by the number of recruits in the division, then summing (i.e., (71*4.00) + (54*3.72) = 484.88). The total test score for the control group in this study was computed in similar fashion (i.e., 22*4.018 = 88.40). The average score for the "untreated" controls in the two recruit divisions was (484.88 - 88.40)/103 = 3.85. This figure actually was lower than the average of 3.94 observed in the previous study.

97
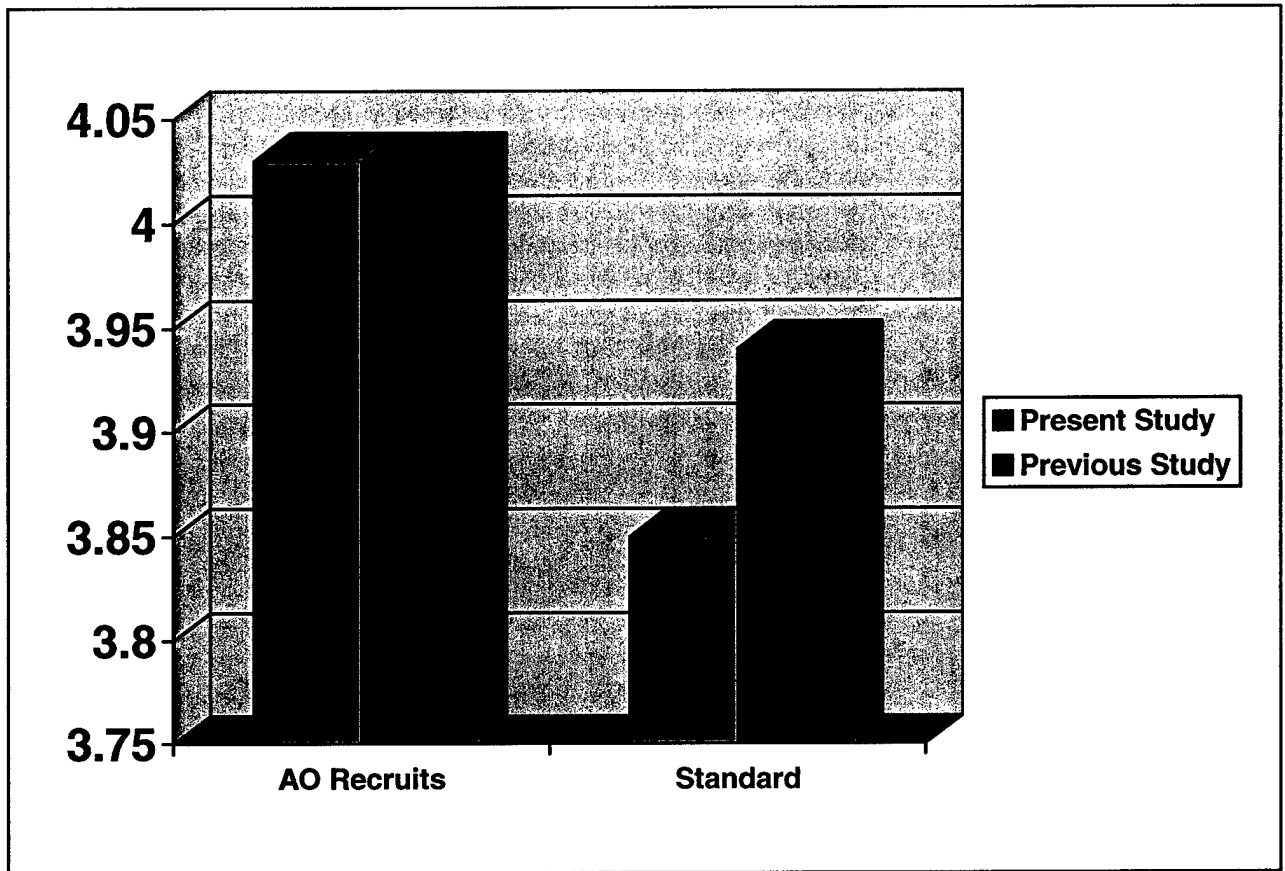
**Figure 9. Raw Academic Test Performance of AO Recruits
and Standard Instruction Recruits**

Note: Sample sizes were Present/AO = 179, Present/Standard = 103, Previous/AO = 423, and Previous/ Standard = 336.

The estimated performance of the "standard training" control group was used for comparisons between the previous study and the present study. Analyses of raw academic test performance, therefore, were undertaken which showed that:

1. The average academic test score for the AO groups in this study was virtually identical to that in the preceding study of AO effects (Present = 4.034; Prior = 4.032).

2. The average academic test score for the control group was slightly lower than that in the previous study (Present = 3.85; Prior = 3.94).

3. The AO-Control group difference was significant in the prior study (Control = 3.94, t_{757} = 2.92, p = .002, one-tailed). Note that the degrees of freedom indicates that the significance test involves a large sample size.

4. The AO-Control group difference was significant in this study. In this study, the average AO recruit who received scored 4.03 (SE = .03) on the academic test. The 95% confidence interval of 3.97 to 4.10 for the AO treatment group. The difference was more pronounced in the present study given the lower average untreated control group score (d = .18, t_{179} = 5.97, p < .001, one-

98

tailed). The comparison still would have been significant if the untreated controls from the prior study had been the point of reference (d = .10, t_{179} = 3.16, p < .001, one-tailed).

The net result of these comparisons was that the difference between the AO treatment groups and untreated controls has been closely replicated across two studies. The differences were statistically significant (p < .002) within each study. The pooled probability for the combined results would be significant at a much higher level. Thus, expressed in terms of the raw academic test score with standard training as the reference point, the AO reliably improved performance. The initial instruction period provided to the control group in this study simply represents another form of Advanced Organizer.

6. Conclusions from Second Data Collection Effort, Year Three (Task 6)

The data analyses reported in this section attempted to replicate the prior findings regarding the impact of AO on performance in overall firefighting training. The analyses attempted to extend those prior observations by testing the hypothesis that the final performance differences reflected no more than the effects of initial learning resulting directly from the exposure to the AO. The effects of differences in AO design also are examined and provided the basis for a number of inferences about the impact of the AO.

- **A course overview helps with final academic performance.** Both the multimedia AO and the classroom/instructor AO improved performance significantly over standard instruction. This finding supports the premise that to overcome difficulties with the content of technical material, represented by the academic component of firefighting, not only do students need analogies they can understand, they need an introduction to the information in a way that will help them bridge the gap between what they know coming into firefighting training and what they need to know to pass the course.

- **The format of the AO may be important in the short term, but not over the full duration of the course.** The multimedia AO produced significantly better performance on the test immediately following the treatment than the classroom/instructor AO. However, the recruits given the classroom/instructor AO "caught up" by the end of the course and they did as well on the final firefighting test as the multimedia AO group. However, even though the classroom/instructor and multimedia AOs had the same end result on performance, the multimedia AO may be more efficient. The multimedia AO significantly improved immediate performance relative to the control, which indicates there was less learning for the AO recruits during the subsequent formal instruction. This inference follows from the equivalent level of final performance. If final performance reflects an asymptote determined by the nature of the material and the abilities of the student, the initial advantage to the multimedia AO suggests it approached this asymptote more rapidly than the classroom/instructor AO. Both methods appear to raise the asymptote, based on comparison to the standard instruction controls. Perhaps a series of multimedia instructional tools would significantly shorten the total required instructional time. This question cannot be resolved in the present design for two reasons. First, more interim measures of learning would need to be developed to fully define the learning curve. Second, additional multimedia instructional tools would need to be developed to determine whether this approach was superior to standard instruction after the initial introduction.

- **Instructional strategies had some impact on the effectiveness of the AO.** The combination of concept map and student-selected questions provided the most opportunity for the student to just "wander around" through the information. Recruits who were in the concept map/student-selected AO strategy did not perform as well on the academic firefighting test. This indicates that it may

be important to avoid giving the recruits too much freedom. This finding should not be given too much weight until it is replicated.

- **Learning style may be important if trainees are forced to view all questions and answers.** The overall pattern of differences was clear. Forcing the student to listen to all the questions and their answers improved performance for field-dependent recruits, but tended to impair performance for field-independent recruits. This finding supports previous studies that showed that field-independent learners do not adhere as well to externally imposed structures (McGee, 1979). They are more capable of developing their own internal referents and they do not need external referents, such as the AO questions, to process information. On the other hand, field-dependent learners were shown to do better with an externally imposed framework, and external referents (McGee, 1979).

- **Affect of Gender.** The study supported previous research findings that showed that females are more likely to be field-dependent learners, and males are more likely to be field-independent (Witkin, 1981). However, exploration of gender issues was not critical to the effectiveness of the AO, and learning style effects were the same for females and males.

- **The AO effects were nontrivial despite what appears to be a modest absolute magnitude.** The combined weighted average of the difference between the AO and the standard instruction control group from this study and the previous study is 0.114 points on the final academic test. The regression analysis in this study indicates that one point on the AFQT is associated with an expected improvement of 0.0107 points on the final academic test. The benefit of the AO, therefore, was equivalent to adding 10.65 points to each recruit's AFQT. AFQT is a reasonably good measure of general intelligence (psychometric 'g') with a standard deviation of 19.04 in the present sample. The AO performance benefit, therefore, was equivalent to 0.559 standard deviations on an IQ test. If the test were transformed to have the accepted SD = 15 for IQ tests, the result would be about 8.4 IQ points. Raising the 'effective IQ' of the average recruit by 8.4 points may not seem like much, but it is equivalent to moving the average recruit up to the 71st percentile on a standardized intelligence test. Taking a recruit with an AFQT that placed him or her at the 25th percentile, the 'added IQ' would move that recruit up to the equivalent of an approximately 46th percentile recruit who had not received the AO. Following this line of reasoning, even if the AO were not implemented for all recruits, having lower ability recruits go through it could significantly improve the overall performance in firefighting training.

III. KEY RESEARCH ACCOMPLISHMENTS

The following are the key research accomplishments emanating from this research.

- Conducted literature reviews of research related to cognitive learning styles; information processing theories; learner control; intelligent tutoring systems; gender differences in learning; stress, role ambiguity and role conflict.

- Analyzed training requirements by conducting a task and skill analysis, training needs assessment, and firefighting course content analysis at the Recruit Training Center.

- Developed and validated a stress evaluation instrument, two instructional treatment interventions, and a learning styles pretest.

- Implemented the instructional interventions and collected data on their effectiveness at the Recruit Training Center. A total of 1,235 recruits participated in the data collection effort.

- Analyzed the data and prepared two annual reports and one final report of findings.

IV. REPORTABLE OUTCOMES

A paper addressing this study will be presented at the Interservice/Industry Training Simulation and Education Conference to be held in Orlando, Florida from November 29 - December 2, 1999. The paper will be published in the conference proceedings. A copy of the paper can be found in Appendix D.

V. CONCLUSIONS

Summary of Results

The work conducted during the past three years included implementation of two instructional treatments and an initial evaluation of their effectiveness. In summary, the initial results indicate that the three null hypotheses postulated for this study were disproved.

- **First Null Hypothesis**

  Providing women with models of effective performance will not reduce stress and will not improve job performance.

  Findings are contrary to this hypothesis. The results indicate that both the Advanced Organizer and Role Model treatments have positive effects, but the two treatments produce effects in different areas. The RM treatment had some positive effects on distress and efficacy. Female recruits were found to feel better about themselves and had a decrease in negative mood. The RM treatment focused on the behaviors required in actual firefighting, and these behaviors and the related group dynamics are

101

fundamental to the psychological constructs of role ambiguity and role conflict. The RM treatment fostered self-efficacy by providing models of people who succeeded in firefighting training.

- **Second Null Hypothesis**

Providing women with aids to structure the learning process will not reduce stress and will not improve job performance.

Findings disproved this hypothesis. The AO treatment did improve learning for both males and females by providing a structure for acquiring and storing technical material. This treatment is linked to stress only indirectly in that stress is reduced when greater knowledge reduces role ambiguity and role conflict and improves the person's sense of self-efficacy. However, stress apparently is not important in the processes linking the AO treatment directly to improved learning. While female recruits were projected to have higher scores in academic tests because of their higher Armed Forces Qualification Test (AFQT) levels, they scored lower than their male counterparts. Academic performance deficiencies for females who completed the AO were reduced. Male scores also improved for the AO participants, albeit not as dramatically as the scores of the females.

- **Third Null Hypothesis**

Providing men and women with models of effective performance will not result in male/female attitudinal changes.

The stress indicators suggested that the RM intervention was found to have a positive effect on female efficacy and negative mood. However, the effect of the RM on the attitude of male recruits toward female recruits could not be assessed because a suitable behavioral criterion representing individual and/or team effectiveness in the hands-on element of firefighting training was not available. These results imply that the RM treatment holds promise of being an effective procedure given an appropriate criterion. The corresponding team performance criteria and measurements necessary to test this hypothesis were not developed at the RTC because of procedural and resource constraints.

Recommendations for Future Work

The recommendations for future work aimed at continuing to address the research topic in this study are as follows.

1. Replicate the RM treatment using a course in which behavioral variables can be demonstrated and measured.

In this study, the RM treatment appeared to be effective; however, we were unable to directly measure its impact. The female recruits who completed the RM treatment had an improved sense of efficacy and a better mood at the end of firefighting training. They felt that they performed better than females who did not complete the RM. Recruits' spontaneous comments at the end of firefighting training revealed that they felt the RM treatment was useful. The study was undertaken with the belief that suitable hands-on firefighting measures would be available to assess individual differences in performance. However, these measures were not developed. For this reason, the effects of the RM treatment on behavioral variables could not be demonstrated in the firefighting training context.

The RM treatment should be replicated on a course where it is possible to directly demonstrate that the treatment has a positive effect on outcomes that are central objectives of the training. The stress effects analyzed in this study provide reason to believe that the RM treatment can affect training outcome

criteria, but sound behavioral assessments of leadership and/or team effectiveness are required to extend this work. Such measures could be developed, but even then the highly structured, carefully controlled setting of firefighting training might be a poor place to examine these effects. Recruit comments suggest that other situations, such as Service Week, provide more extensive leadership and team performance opportunities. Some recruits commented that they felt this treatment would have helped them during that time period. However, developing this line of inquiry was beyond the scope of available resources for this project. The RM treatment should be given consideration for further research.

2. Determine the point within the Recruit Training curriculum where it would be most appropriate to teach the students "how to learn."

At the onset of boot camp, recruits are undergoing culture shock. During their initial academic sessions, they sit in classrooms like automatons. This phase gradually wears off as the recruits progress in their training. The AO treatment, which was a tool that helped the students learn firefighting training technical material, was proven to be effective. However, it was presented to the recruits during the fifth week of boot camp training, at the very beginning of the firefighting training course, which may not be the optimal time to introduce educational tools. Are the tools most useful early in training, in the middle of training when the recruits' culture shock wears off, or at the end of training as the recruits are preparing for initial skill training? Further research should be conducted to study the effects of a learning tool such as the AO at various times during RTC training. This research would help to determine the optimum point at which this type of learning tool is most valuable.

3. Develop interim measures of learning and additional instructional multimedia tools to determine if the AO approach is superior to standard instruction after the initial introduction of the content.

The results from the study revealed that the *format* of an AO (interactive multimedia versus traditional classroom instruction) was important in the short term but not over the full duration of the course. The multimedia AO produced significantly better performance on the test immediately following the treatment than the classroom/instructor AO. However, the recruits given the classroom/instructor AO "caught up" by the end of the course and they did as well on the final firefighting test as the multimedia AO group. Even though the classroom/instructor and multimedia AOs had the same end result on performance, the multimedia AO may be more efficient. The multimedia AO significantly improved immediate performance relative to the control, which indicates there was less learning for the AO recruits during the subsequent formal instruction. This inference follows from the equivalent level of final performance. If final performance reflects an asymptote determined by the nature of the material and the abilities of the student, the initial advantage to the multimedia AO suggests it approached this asymptote more rapidly than the classroom/instructor AO. Both methods appear to raise the asymptote, based on comparison to the standard instruction controls. Perhaps a series of multimedia instructional tools would significantly shorten the total required instructional time.

This question requires further research. Interim measures of learning need to be developed to fully define the learning curve, and additional multimedia instructional tools need to be developed to determine whether this approach was superior to standard instruction after the initial introduction.

4. Replicate the study to determine if recruits should be provided only with instructional tools that are highly structured.

The instructional strategy in the AO that combined the concept map approach and student-selected questions provided the most opportunity for the student to "go anywhere and do anything" in the multimedia instruction. This AO strategy provided a highly individualized instructional approach. The data revealed that the students in this group did not perform as well on the academic firefighting test as

the students who were in one of the more structured AO strategies. This finding indicates that it may be important to avoid giving this particular target audience too much "freedom" during instruction. Left to their own devices, this target audience may learn the facts, rules, methods, and details, but they may not understand how or why they fit together. They may be gaining only a sparse mental model of the material and their recall of what they learned may be guided arbitrarily and accidentally.

The reason for this may be that from the very start of boot camp training, recruits are not provided with individualized training. Every recruit, regardless of education, ethnicity, gender, or age, must perform the same tasks, in the same way, to the same standards of performance. The recruits are not segregated into different groups based on their educational sophistication or ability to learn the tasks and skills. The military environment at boot camp forces all the of the recruits to adopt a structured, serialist approach to learning. Further research must be conducted to determine if encouraging the use of personal learning strategies improves overall learning, retention, and transfer of information to skills mastery.

5. Conduct a study which focuses on gender differences in learning using conversational techniques.

This study supported previous research findings that showed that females are more likely to be field-dependent learners while males are more likely to be field-independent learners. The study also revealed that learning style and learning strategy are related. However, the data revealed that gender was not critical to the effectiveness of the AO, and learning style effects were the same for females and males.

A second study must be conducted which focuses solely on male/female differences in learning style. Based on the findings from this study, we recommend that conversational theory (Pask, 1976) be the basis for the additional research. Conversational techniques allow mental activities to be described in terms of dialog and behavior. Whether provided by an instructor in a classroom or through interactive multimedia courseware, students would be forced to make their learning strategy explicit via their verbal and behavioral explanation and derivation of the concepts or topics. As in the reported study, the students' learning style is characterized as holist (field-dependent) or serialist (field-independent). Having adapted a learning strategy, research has shown that students do not relinquish it, even if they cannot successfully execute it. Sharp strategic distinctions occur because the students become "locked into" one strategy to the exclusion of others (Pask, 1976).

Further research needs to be conducted to continue to explore the hypothesis that matching a teaching strategy to learning style will result in students learning more quickly and retaining the information for longer periods of time. In this study we showed that a "mismatched condition" led to inferior performance and a proven failure on the part of the recruits to comprehend the principles underlying the subject matter. Students who were tested and shown to be field-dependent learners performed better in the strategy where they were provided with all the questions and answers available in the tutoring system. On the other hand, field-independent learners did not perform as well when forced to use that strategy. This finding supported Pask's (1976) research that matching learning style with learning strategy improved students' performance. The finding also supported McGee's (1979) research that showed that field-independent learners do not adhere as well to externally imposed structures, while field-dependent learners were shown to do better with an externally imposed framework.

Evaluation of Knowledge in Terms of Scientific Product

The knowledge gained from this study has provided a documented scientific framework within which to view and measure stress in the recruit training arena. The knowledge gained from this study has also provided a documented scientific framework which has allowed us to expand educational concepts by integrating the current works of numerous learning theorists.

The strongest underlying assumption of this study was that the female recruits were different from male recruits in their use of coping, communication and learning strategies. It was posited that these differences contributed significantly to various stress levels among recruits. Specifically, it was felt that placing females in a male-dominated environment such as the Navy (13.7% female, 86.3% male) would increase the stress level of females such that their learning capabilities would be significantly reduced. These assumptions were grounded in the belief that women were not performing as well as men in technical subjects such as firefighting training. The attrition rate for women was also higher than that of men. Reports of misconduct between male instructors with female recruits brought into question the entire military training environment. The study assumed that all these factors contributed to a high level of stress for female recruits. However, findings indicate that females and males generally reported comparable stress levels. The only significant difference between genders was for females to report lower role conflict after firefighting training. While this finding is contrary to popular perception, it is consistent with prior research (Martocchio and O'Leary, 1989). *The general belief that females are more susceptible to stress is, for boot camp, a misleading stereotype.* Recognition of this fallacy allowed redirection of the study to focus on an enhanced understanding of strategies to improve learning within a military environment.

Accordingly, the focus of the study addressed disproving the second hypothesis, that providing women with aids to structure the learning process would not reduce stress and would not improve job performance. The AO provided a range of interaction between student and computer that reflected field-dependent and field-independent learning concepts. An outcome of its use was that students were taught sophisticated principles of learning in a very simple non-technical manner. It keyed integral concepts to the student's prior knowledge, and provided a course overview in such a manner that it focused the learners to either a global or a more specific view of the course material—via concept mapping or linear strategies.

The results were impressive and not unexpected. Initial findings during Year Two were in keeping with the work of Williams (1993) that showed that students with high ability benefited more from learner control strategies than other types of students. The Year Three results indicated that for students with lower AFQT scores, the effects of the AO were significantly reduced. However, even though the AO effect was small in terms of absolute magnitude, AO effects were significant. In essence, use of the AO added 10.65 points to the AFQT. More generally stated, AO raised the IQ by an estimated 8.4 IQ points.

The inferences from these findings have the following implications for Navy pedagogy:

- A course overview helps with final academic performance. The implication of this finding relates to "training the trainer." Navy instructors need to be educated about learner styles and strategies to address those styles while still observing the military decorum of boot camp. This outcome strongly evidences Ausubel's (1978) contention that the single most important factor influencing learning is the experiential background of the student.

- AO format is important. The multimedia AO produced better performance on the test immediately following the treatment. Implications are that in some topic areas that lend themselves to technology, technology increases the efficiency of the instruction. In the military training arena, reducing class length leads to cost savings. In boot camp, reducing class length increases capacity, i.e., more students can be trained in the same amount of time.

- Raising the effective IQ score of lower-ability recruits improves academic performance. With implementation of the AO, even the lowest performing recruit's IQ was elevated from the 25th percentile to the 46th. It follows that a re-engineering of remedial training to use strategies such as the AO will benefit the students who need the most academic help.

- The issues of instructional philosophy and design were affected. In this study, various learning theories, instructional approaches and teaching techniques were reviewed to address the academic deficiencies of the target population. We synthesized the elements of major instructional design theories and research findings into a single integrated model. The approach to the study, and the development of the educational tools, demonstrated that this synergistic model of curriculum development was very effective. The research findings were used as a basis for deriving the new model based on the premise that to overcome difficulties with the content of technical material represented by the academic component of firefighting training, not only do recruits need analogies they can understand, they need an introduction to the information in a way that complements their unique learning styles.

Recruit training and initial skills training have been undergoing re-engineering in the Navy to improve instruction; however, the efforts have been eclectic. If the Navy were to redefine its educational philosophy and instructional models using a synergistic all-encompassing theoretical framework, such as the one implemented during this study, a more effective instructional approach would be enforced. A consequence of this approach is that the Curriculum Instruction Standards Officers, who oversee all curriculum development initiatives in the Navy, are provided with new guidance. The data yielded by this study, and the instructional approach which was implemented, have affected Navy policy decisions and instructional philosophies. Future Navy training programs now have a sound empirical base, built on concrete data, and they can be implemented using a comprehensive educational approach.

VI. REFERENCES

Ackerman, P.L., Sternberg, R.J., Glaser, R., Eds. (1989). Learning and Individual Differences: Advances in Theory and Research. New York: W.H. Freeman & Co.

Ash, B. (1986). Identifying learning styles and matching strategies for teaching and learning. (ERIC Document Reproduction Service No. 270142).

Ausubel, D.P., Novak, J.D., and Hanesian, H. (1978). Educational Psychology: a Cognitive View (Second Edition). New York: Holt, Rinehart and Winston.

Bailey, S.S., Johnston, J.H., Smith-Jentsch, K.A., Gonos, G., & Cannon-Bowers, J.A. (1995). Guidelines for facilitating shipboard team training. Proceedings of the 17th annual Interservice/Industry Training Systems and Education Conference, Washington, DC, 360-369.

Bentler, P.M. (1989). EQS structural equations program manual. Los Angeles: BMDP Statistical Software, Inc.

Bentler, P.M., & Bonett, D.G. (1980). Significance tests and the goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Bollen, K.A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. Psychological Bulletin, 110, 305-314.

Brookfield, S.D. (1986). Understanding and Facilitating Adult Learning. San Francisco: Jossey-Bass, Inc.

Cannon-Bowers, J.A. & Salas, E. (in press). A framework for developing team performance measures in training. To appear in M.T. Brannick, E. Salas, & C. Prince (Eds.), Team Performance Assessment and Measurement: Theory, Methods, and Applications, Hillside, NJ: Lawrence Erlbaum.

Cannon-Bowers, J.A. & Salas, E. (1990). Teamwork competencies: The interaction of team member knowledge, skills, and attitudes. To appear in H.F. O'Neil (Ed.), Workforce Readiness: Competencies and Assessment, Hillside, NJ: Lawrence Erlbaum.

Cannon-Bowers, J.A., Salas, E., & Baker, C.V. (1991). Do you see what I see? Instructional strategies for tactical decision making teams. Proceedings of the 13th annual Interservice/Industry Training Systems Conference, Washington, DC 214-220.

Cannon-Bowers, J.A. Salas, E., & Converse, S.A. (1993). Shared mental models in expert team decision making. In N.J. Castellan, Jr. (Ed.). Individual and Group Decision Making: Current Issues, pp. 221-246, Hillsdale, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J.A., Salas, E., Duncan, P., & Halley, E.J. (1994). Application of multi-media technology to training for knowledge-rich systems. Proceedings of the 16th annual Interservice/Industrial Training Systems Conference. Washington, DC, 6-11.

Caplan, R.D., Cobb, S., French, J.R.P., Jr., Harrison, R.V., & Pinneau, S.R., Jr. (1980). Job demands and worker health. Cincinnati: National Institute for Occupational Safety and Health.

Carroll, J.B. (1993). Human cognitive abilities: A survey of factor-analytic studies. NY: Cambridge.

Castenada, A., Ramirez, M. III, & Herold, P. (1972) Culturally democratic learning environments: A cognitive styles approach. Multi-lingual assessment project. Riverside, CA: Systems and Evaluation in Education.

Cates, W.M. (1992). Fifteen principles for designing more effective instructional hypermedia/multimedia products. Educational Technology, 32(12), 5-11.

Chung, J., & Reigeluth, C.M. (1992). Instructional prescriptions for learner control. Educational Technology, 32(10), 14-20.

Clark, R.E. (1983). Research on student process during computer-based instruction. Journal of Instructional Development, 7(3), 2-5.

Clinchy, B.M. (1995). A connected approach to the teaching of developmental psychology. Teaching of Psychology, 22(2), 100-104.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). NY: Academic Press.

Cota, A.A., Longman, R.S., Holden, R.R., Fekken, G.C., & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. Educational and Psychological Measurement, 53, 585-596.

Daniels, H.L. (1996). Interaction of cognitive style and learner control of presentation mode in a hypermedia environment. Doctoral dissertation. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Duncan, P.C., Burns, J.J. & Frey, P.R., Cannon-Bowers, J.A. & Johnston, J.H. (1995). Enhancing team performance in tactical environments: The team model trainer. Proceedings of the 17th Annual Interservice/Industry Training Systems and Education Conference, Washington, DC, 262-271.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. Psychological Bulletin, 116, 429-456.

Fischer, G., Lemke, A.C. & McCall, R. (1990). Towards a system architecture supporting contextualized learning. Education and Artificial Intelligence, 1, 420-425.

Fleishman, E.A., & Quanitance, M.K. (1984). Taxonomies of human performance: The description of human tasks. Orlando: Academic Press.

Frank, B.M. (1986). Cognitive styles and teacher education: Field dependence and areas of specialization among teacher education majors. Journal of Educational Research, 80(1), 19-22.

Frey, D., & Simonson, M. (1994). Assessment of cognitive style to examine student's use of hypermedia within historic costume. Home Economics Research Journal, 21(4), 403-421.

Gallos, J.V. (1995). Gender and silence: Implications of women's ways of knowing. College Teaching, 43(3), 101-105.

Gebhardt, D.L., Jennings, M.C., & Fleishman, E.A. (1981). Factors affecting the reliability of physical ability and effort ratings of Navy tasks (Tech. Rep. R81-1). Washington, DC: Advanced Research Resources Organization.

Glickman, A.S., Zimmer, S. Montero, R.C. Guerette, P.J., Campbell, W.J., Morgan, B.B., Jr., & Salas, E. (1987). The Evolution of Teamwork Skills: An Empirical Assessment with Implications for Training. (Technical Report No. 870016). Orlando, FL: Naval Training Systems Center.

Glymour, C., Schienes, R., Spirtes, P., & Kelly, K. (1987). Discovering causal structure: artificial intelligence, philosophy of science, and statistical modeling. NY: Academic Press.

Gray, G.S. (1987). The effect of sequence control on computer-assisted instruction. Journal of Computer-Based Instruction, 14(2), 54-56.

Gray, John (1992). Men are from Mars, Women are from Venus. New York: Harper, Collins.

Hansen, J.W., (1994). Student cognitive styles in postsecondary technology programs. Journal of Technology Education, 6(2), 1045-1064.

Harris, R.J. (1985). A primer of multivariate statistics (2nd ed.). Orlando, FL: Academic Press.

Hays, W.L. (1963). Statistics for psychologists. New York: Holt, Rinehart, & Winston.

Hsi, S. & Hoadley, C.M. (1994). An interactive multimedia kiosk as a tool for collaborative discourse, reflection, and assessment. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Jick, T.D., & Mitz, L.F. (1985). Sex differences in work stress. Academy of Management Review, 10, 408-420.

Johnston, J.H., Cannon-Bowers, J.A., & Smith-Jentsch, K.A. (1995). Event-based performance measurement system for shipboard command teams. Proceedings of the First International Symposium on Command and Control Research and Technology, Washington, DC, 274-276.

Jonassen, D.H. (1988). Integrating learning strategies into courseware to facilitate deeper processing. In D.H. Jonassen (Ed.), Instructional Designs for Microcomputer Courseware (pp. 151-181). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jonassen, D.H. (1989). Hypertext/Hypermedia. Englewood Cliffs, NJ: Educational Technology Publications.

Jonassen, D.H. (1991). Representing the expert's knowledge in hypertext. Impact Assessment Bulletin, 9(1-2), 93-105.

Jonassen, D.H. & Grabowski, B.L. (1993). Handbook of individual difference. Learning and Instruction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Joreskog, K.G., & Sorbom, D. (1989). LISREL VII (2nd Ed.). Chicago: SPSS, Inc.

Kagan, S. & Zahn, G.L. (1975). Field-dependence and the school achievement gap between Anglo-American and Mexican-American children. Journal of Educational Psychology, 67(5), 643-650.

Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D., & Rosenthal, R.A. (1964). Occupational stress: Studies in role conflict and ambiguity. NY: Wiley.

King, L.A., & King, D.W. (1990). Role conflict and role ambiguity: A critical assessment of construct validity. Psychological Bulletin, 107, 48-64.

Kolb, D.A. (1976). Learning Style Inventory Technical Manual. Boston, MA: McBer and Co.

Kraer, K. & Cannon-Bowers, J. (1995). Measuring knowledge organization as a method for assessing learning during training. Human Factors, 37(4), 804-816.

Kurland, L.C. and Tenney, Y.J. (1988). The development of troubleshooting expertise in radar mechanics. In J. Psotka, L. Massey, & S. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 59-83). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kurland, L.C. and Tenney, Y.J. (1988). Issues in Developing an intelligent tutor for a real-world domain: training in radar mechanics. In J. Psotka, L. Massey, & S. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 119-180). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, W. & Pangaro, P. (1994). Safety and understanding: design and development of intelligent training. Paper presented at the annual meeting of the American Institute of Chemical Engineers.

Lin, X. (1994). Embedding metacognitive cues into hypermedia systems to promote far transfer problem solving. ERIC Document Reproduction Service No. ED 373 735.

Lorr, M., McNair, D.M., & Fisher, S. (1982). Evidence for bipolar mood states. Journal of Personality Assessment, 46, 432-436.

Marchionini, G. (1988). Hypermedia and learning: Freedom and chaos. Educational Technology, 28(11), 8-12.

Martocchio, J.J., & O'Leary, A.M. (1989). Sex differences in occupational stress: A meta-analytic review. Journal of Applied Psychology, 74, 495-501.

Massey, L.D., deBruin, J., & Roberts, B. (1988). A training system for system maintenance. In J. Psotka, L. Massey, & S. Mutter (Eds.), Intelligent Tutoring Systems: Lessons Learned (pp. 368-402). Hillsdale, NJ: Lawrence Erlbaum Associates.

McEneany, J.E. (1990). Do AOs facilitate learning? A review of subsumption theory. Journal of Research and Development in Education, 23(2), 89-96.

McGee, M.G. (1979). Human spatial abilities: Source of sex differences. New York: Praeger.

McIntyre, R.M., Morgan, B.B., Jr., Salas, E., & Glickman, A.S. (1988). Team research in the eighties: Lessons learned. Unpublished manuscript, Naval Training Systems Center, Orlando, FL.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1981). Manual: Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service.

Merrill, M.D. (1984). What is learner control? In C.R. Dills (Ed.), Instructional development: The state of the art, II, (pp. 221-242). ERIC Document Reproduction Service No. ED 298 905.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. Educational Psychologist, 19(2), 59-74.

Mitchell, T.R. (1974). Expectancy model of job satisfaction, occupational performance and effort: A theoretical, methodological, and empirical appraisal. Psychological Bulletin, 81, 1053-1077.

Moore, D.M., Burton, J.K., & Myers, R.J. (in press). Multiple-channel communication: The theoretical and research foundations of multimedia. In D.H. Jonassen (Ed.), Handbook of research on communication and educational technology. Washington, DC: Association for Educational Communications and Technology.

Mulaik, S.A., James, L.R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Muraida, D.J. & Spector, M. (1993). The advanced instructional design advisor. Instructional Science, 21, 239-253.

Nelson, W.A., & Palumbo, D.B. (1992). Learning, instruction, and hypermedia. Journal of Educational Multimedia and Hypermedia, 1(1), 287-299.

Nunnally, J.C. & Bernstein, I.H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Olstad, R.G., Juarez, J.R., Davenport, L.J., & Haury, D.L. (1981). Inhibitors to achievement in science and mathematics by ethnic minorities. (ERIC Document Reproduction Service No. 223 404).

Packard, A.L. (1996). Exploration of a Brunswick learning environment developed to instruct basic statistical concepts. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg.

Pangaro, Paul A. (1987). An Examination and confirmation of a macro theory of conversations through the realization of the protologic lp by microscopic simulation. A thesis submitted for the degree of Doctor of Philosophy, Department of Cybernetics, Brunel University.

Park, O. (1991). Hypermedia: Functional features and research issues. Educational Technology, 31(8), 24-31.

Pask, G. (1976). Styles and strategies of learning. Journal of Educational Psychology, 46, 128-148.

Philbin, M., Meier, E. Huffman, S., & Boverie, P. (1995). A survey of gender and learning styles. Sex Roles: A Journal of Research, 32 (7-8), 485-94.

Ramirez, M. III, & Price-Williams. (1974). Cognitive styles of children of three ethnic groups in the United States. Journal of Cross-Cultural Psychology, 5, 212-219.

Recker, M.M., & Pirolli, P. (1992). Student strategies for learning programming from a computational environment. In C. Frasson, G. Gauthier, & G.I. McCalla (Eds.), Lecture notes in computer science: Intelligent tutoring systems. Berlin: Springer-Verlag. (382-394).

Ree, M.J., & Earles, J.A. (1991). Predicting Training Success. Personnel Psychology, 44, 321-332.

Reigeluth, C.M., & Stein, F.S. (1983). The elaboration theory of instruction. In C.M. Reigeluth (Ed.), Instructional design theories and models (pp. 335-382). Hillsdale, NJ: Lawrence Erlbaum.

Reynolds, D.H., Barnes, J.D., Harris, D.A., & Harris, J.H. (1992). Analysis and clustering of entry-level Navy ratings (Final Report HumRRO FR-PRD-92-20). Alexandria, VA: Human Resources Research Organization.

Rideout, C.A. & Richardson, S.A. (1989). A teambuilding model: Appreciating differences using the Myers-Briggs Type Indicator with developmental theory. Journal of Counseling and Development, 67, 529-533.

Rollock, D. (1922). Field dependence/independence and learning condition: An exploratory study of style vs. ability. Perceptual and Motor Skills, 74, 807-818.

Rosenthal, R., & Rosnow, R.L. (1984). Essential of behavioral research: methods and data analysis. NY: McGraw-Hill.

Ross, S.M. (1984). Matching the lesson to the student: Alternative adaptive designs for individualized learning systems. Journal of Computer-Based Instruction, (11)2, 42-48.

Rowland, P. & Stuessy, C.L. (1988). Matching mode of CAI to cognitive style: An exploratory study. Journal of Computers in Mathematics and Science Teaching, (7)4, 36-40.

Russell, J.A., & Steiger, J.H. (1982). The structure in persons' implicit taxonomy of emotions. Journal of Research in Personality, 16, 447-469.

Ryman, D.H., Biersner, R.J., & LaRocco, J.M. (1974). Reliabilities and validities of the mood questionnaire. Psychological Reports, 35, 479-484.

Salas, E., Bowers, C.A., & Cannon-Bowers, J.A. (1995). Military team research: 10 years of progress. Military Psychology, 7(2), 55-75.

Saunders, D.R. (1956). Moderator variables in prediction. Educational and Psychological Measurement, 16, 209-222.

Shaughnessy, J.J. & Zechmeister, E.B. (1985). Research methods in psychology. New York: Knopf.

Sherer, M., Maddux, J.E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R.W. (1982). The Self-Efficacy Scale: Construction and Validation. Psychological Reports, 51, 663-671.

Shlechter, T. (1986). An Examination of the Research Evidence for Computer-Based Instruction in Military Training. Technical Report 722.

Slater (1997). Personal conversation.

Small, R.V. & Grabowski, B.L. (1992). Exploratory study of information seeking behaviors and learning with hypermedia information systems.

SPSS, Inc. (1990). SPSS reference guide. Chicago: SPSS, Inc.

Steinberg, E.R. (1989). Review of student control in computer-assisted instruction. Journal of Computer-Based Instruction, 16(4), 117-124.

Tannen, Deborah (1990). You Just Don't Understand. New York: William Morrow and Company, Inc.

Tannen, Deborah (1994). Talking from 9 to 5. New York: Avon Books.

Theil, J.P. (1984). Successful self-directed learners' learning styles. Proceedings of the Adult Education Research Conference, No. 25. Raleigh: North Carolina State University.

Tisdell, E.J. (1993). Feminism and adult learning: Power, pedagogy, and praxis. New Directions for Adult and Continuing Education, 5, 91-103.

Tripp, S.D., & Roby, W. (1990). Orientation and disorientation in a hypertext lexicon. Journal of Computer-Based Instruction, 17(4), 120-124.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Vickers, R.R., Jr., & Hervig, L.K. (1987). Effects of response style on the polarity and validity of two-dimensional mood models. Psychological Reports, 61, 247-258.

Vickers, R.R., Jr., & Ryman, D.H. (1980). Development of a questionnaire to measure psychological stress and related concepts in the context of Marine Corps Basic training setting (Tech. Rep. 80-12). San Diego, CA: Naval Health Research Center.

Volpe, C.E., Cannon-Bowers, J.A., & Salas, E. (1996). The impact of cross-training on team functioning: An empirical investigation, Human Factors, 38(1), 87-100.

Wallick, M.T., Vickers, R R., Jr., & Ryman, D.H. (1982). Revision of a questionnaire to measure stress and related aspects of basic training (Tech. Rep. 82-19). San Diego, CA: Naval Health Research Center.

Watson, D., Clark, L.A., & Tellegen, A. (1985). A. Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. Psychological Bulletin, 98, 219-235.

Weller, H.G., Repman, J., & Rooze, G.E. (1994). The relationship of learning, behavior, and cognitive styles in hypermedia-based instruction: Implications for design of HBI. Computers in Schools, 10(3-4), 401-420.

Williams, M.D. (1993). A comprehensive review of learner-control: The role of learner characteristics. Paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, LA.

Winer, B.J. (1962). Statistical principles in experimental design (pp. 174-178). NY: McGraw-Hill.

Winfred, A. Young, B., Jordan, J. & Shebilske, W.L. (1996). Effectiveness of individual and dyadic training protocols: The influence of trainee interaction anxiety. Human Factors, 38(1), 79-86.

Witkin, H.A. (1950). Individual differences in ease of perception in embedded figures. Journal of Personality, 19, 1-15.

Witkin, H.A. & Goodenough, D.R. (1981). Cognitive styles: Essence and Origins. New York: International Universities.

Witkin, H.A., Moor, C.A., Oltman, P.K., Goodenough, D.R., Friedman, F., Owen, D.R., & Raskin, E. (1977). Role of field-dependent and field-independent cognitive styles in academic evolution: A longitudinal study. Journal of Educational Psychology, 69(3), 197-211.

Witkin, H.A., Oltman, P.K., Raskin, E. & Karp, S.A. (1971). A Manual for the Embedded Figures Tests, Consulting Psychologists Press.

Zedeck, S. (1971). Problems with the use of moderator variables. Psychological Bulletin, 76, 295-310.

VII. BIBLIOGRAPHY OF PUBLICATION AND PERSONNEL RECEIVING PAY FROM THE RESEARCH EFFORT

The publication resulting from this research effort is as follows:

Golas, K.C., Bartoli, C.S., Miller, S., & Idar, I. (1999). Research and Development of Intelligent Tutoring Strategies for U.S. Naval Recruits. Proceedings of the 1999 Interservice/Industry Training Simulation and Education Conference. Orlando, FL.

The personnel receiving pay from the research effort are as follows.

Southwest Research Institute

Katharine Golas, Principal Investigator
Claire Bartoli, Project Manager
Sarina Goodman, Instructional Designer
Sherri Miller, Programmer
Linda Estes, Secretary
Christine Reed, Secretary
Tammy Swanson, Programmer
Kevin Fiedler, Programmer
Elizabeth Branch, Editor
Anna Longoria, Administrative Assistant

U.S. Navy

Ross Vickers, Statistician

APPENDIX A

SCREEN DISPLAYS FROM THE TWO INTERVENTIONS

Advanced Organizer

INTRODUCTION TO FIREFIGHTING

Please fill in the following information:

SSAN (###-##-####)   123-34-4567

Last Name   DOE

First Name   JANE

Gender (M or F)   F

Division   123

QUIT

Student Tracking - Data Collection

Advanced Organizer

INTRODUCTION TO FIREFIGHTING TRAINING

To begin, press any key.

Advanced Organizer

NAVIGATION HELP

Click on any of the navigation buttons for help information.

PAUSE ❙❙

REPLAY ▶

Click **OK** when you're ready to leave "HELP".

| QUIT | HELP | TEST | TOPICS | BACK | ? | FORWARD |

Advanced Organizer

9. Classes of Fire and Fire Extinguishing Agents & Systems

QUESTIONS?

1. I am not a DC man. Why is this information important to me?

2. Is the weight of the equipment a problem for females?

3. What happens if I use the wrong extinguishing agent or system?

4. After recruit training, will I receive further training on these systems?

5. What exactly do I need to remember?

Capability to Ask Questions

Advanced Organizer

1. Nature of Fire

Heat

PAUSE ❚❚   REPLAY ▶

Fire Tetrahedron:

Fuel
Oxygen
Heat
Uninhibited Chain Reaction

QUIT | HELP | TEST | TOPICS | ? | FORWARD

Video

Advanced Organizer

4. Compartment Identification
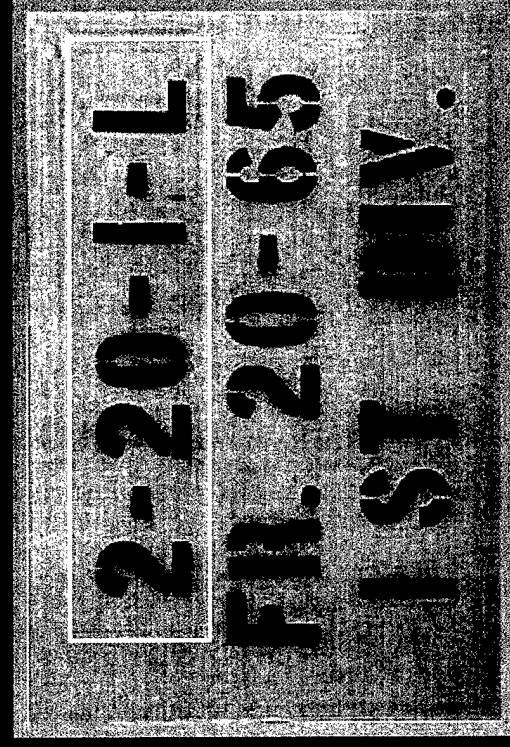
Compartment's address or ID.

2-20-I-L
FL 20-65
1 ST DIV.

QUIT | HELP | TEST | TOPICS | BACK | ? | FORWARD

Photographs and Graphics

Advanced Organizer

7. Protective Clothing

Helmet
Antiflash Hood
OBA
Gloves
Coveralls
Boots
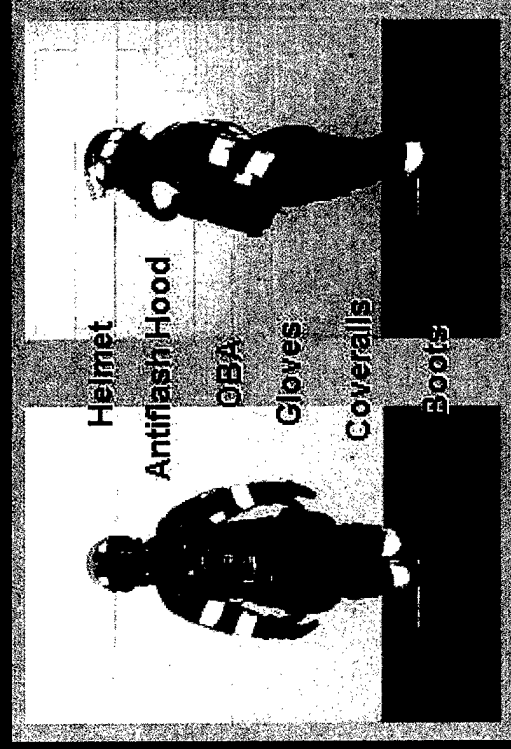
QUIT | HELP | TEST | TOPICS | BACK | ? | FORWARD

Student Interaction

Advanced Organizer

10. Dewatering Equipment

HIGHER PRESSURE

LOWER PRESSURE

HIGHER PRESSURE

REPLAY ▶
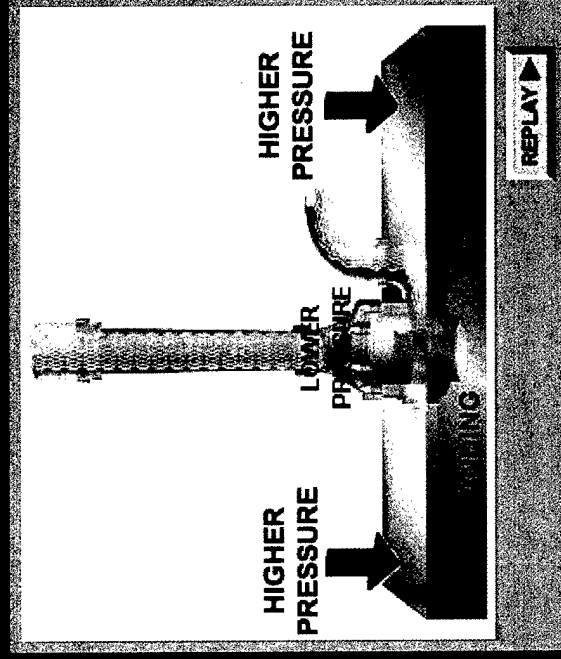
QUIT | HELP | TEST | TOPICS | BACK | ? | FORWARD

Animation

Advanced Organizer

4. Compartment Identification

2-75-2-L
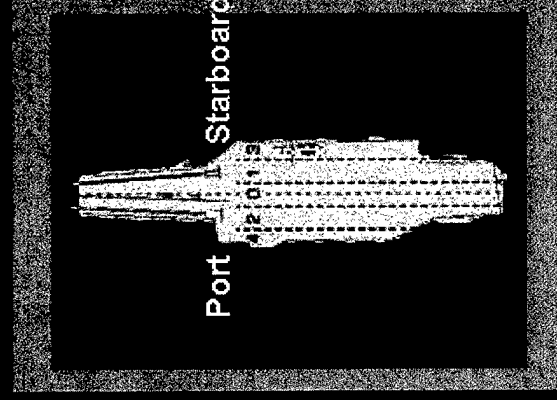
Port   Starboard

PESO

Port - Even - Starboard - Odd

QUIT   HELP   TEST   TOPICS   BACK   ?   FORWARD

Audio with Support Text

Advanced Organizer

5. Material Conditions of Readiness

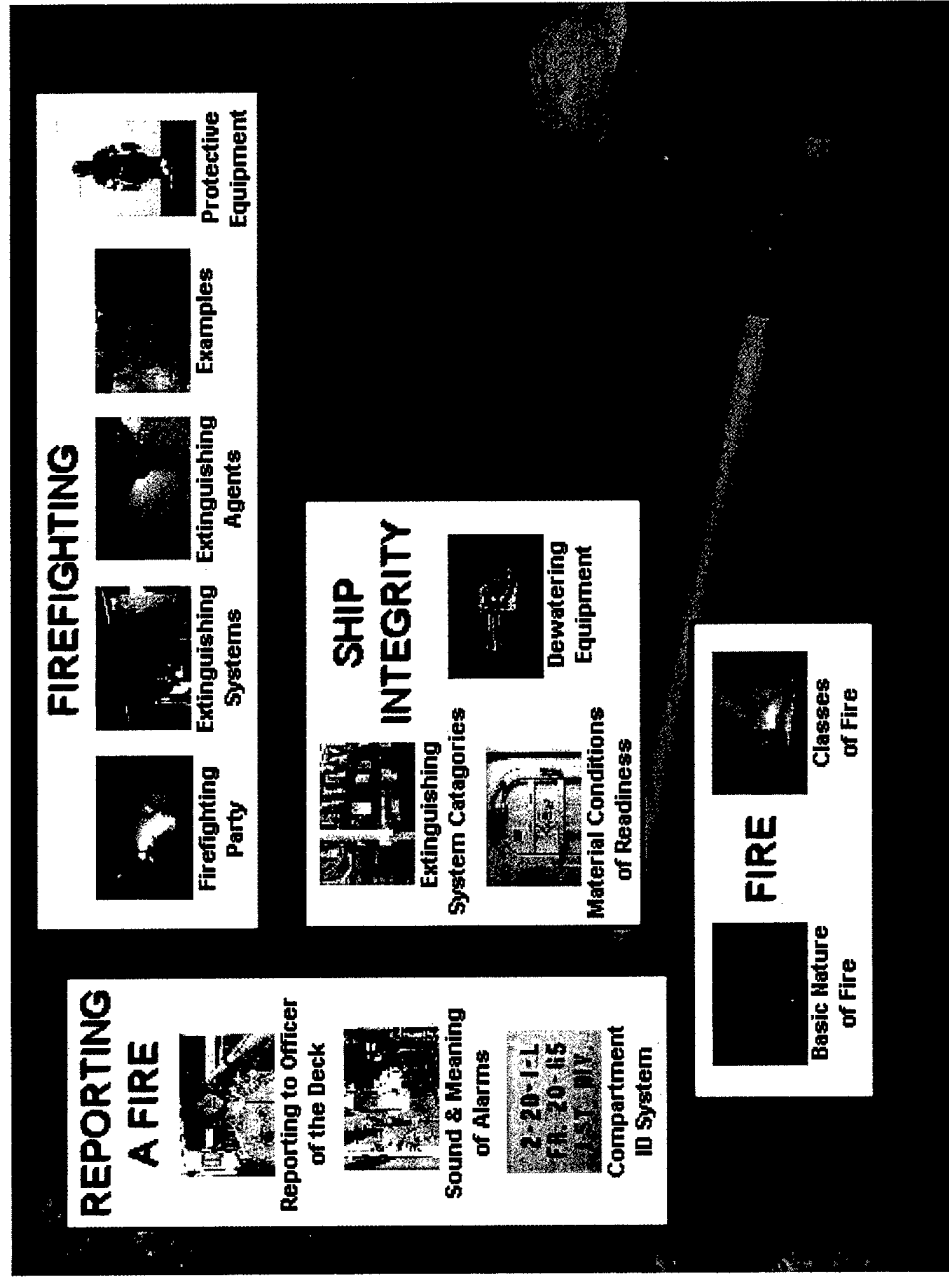### FIREFIGHTING TOPICS

1. Nature of Fire
2. How to Report a Fire
3. Response to Fire Alarm
4. Compartment Identification
5. Material Conditions of Readiness
6. Breathing Devices
7. Protective Clothing
8. Performing in a Firefighting Party
9. Classes of Fire and Fire Extinguishing Agents & Systems
10. Dewatering Equipment

[OK]

Topics Covered - Yet to Cover

Advanced Organizer

FIREFIGHTING

Firefighting Party

Extinguishing Systems

Extinguishing Agents

Examples

Protective Equipment

REPORTING A FIRE

Reporting to Officer of the Deck

Sound & Meaning of Alarms

2-20-1-1

Compartment ID System

SHIP INTEGRITY

Extinguishing System Catagories

Material Conditions of Readiness

Dewatering Equipment

FIRE

Basic Nature of Fire

Classes of Fire

Concept Map

Advanced Organizer

Test

Question 1: What type of information must be given when reporting a fire?

A Type of emergency

B Location of emergency

C Your name

D Your rate or rank

E Your telephone number
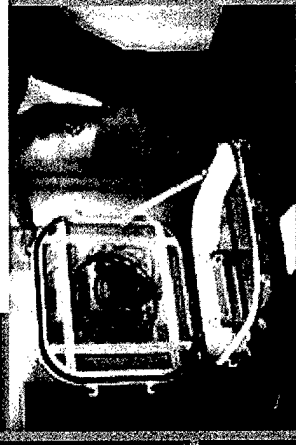
F Your immediate supervisor's name and rank

QUIT

FORWARD

Test - Multiple Answers

Advanced Organizer

Test

Question 5: After hearing the announcement, you immediately know that you must close all of your assigned fittings. Click on the fittings that you should close. To hear the audio again, click on the Replay button.
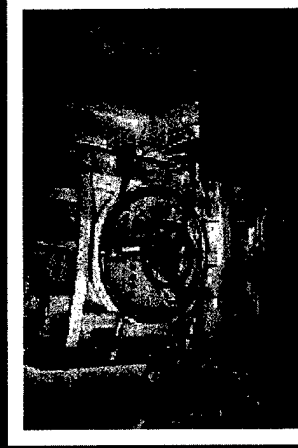
FORWARD

REPLAY ▶

QUIT

Audio Prompt - Action Response

Advanced Organizer

Test

Question 8: Click on the appropriate clothing and equipment to fight a fire.

OBA

Eductor

Gloves

Safety goggles

Hard-soled, steel-toed boots

Helmet

Coveralls

EEBD

Standard-issue boots

Anti-flash hood

QUIT

FORWARD

Test - List

Advanced Organizer

Test

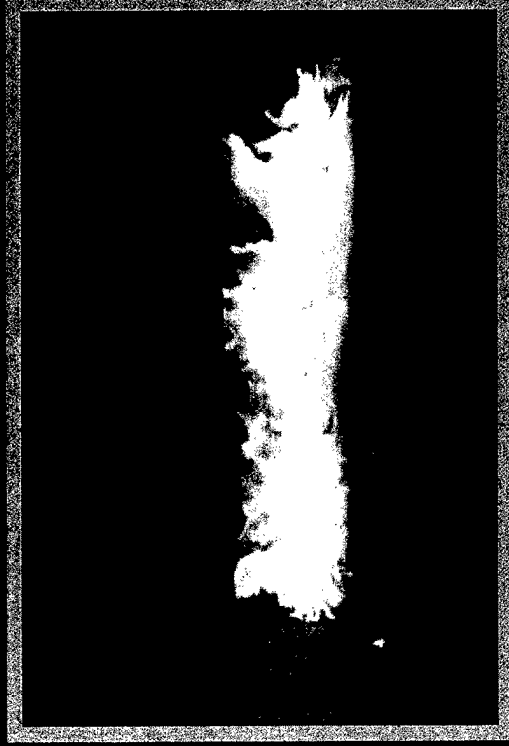Question 2: While waiting for the firefighting team to arrive, you should:

A   Attempt to extinguish the fire in all circumstances

B   Attempt to extinguish only small fires

C   Put on a OBA

D   Sound general alarm

QUIT   FORWARD

Test - Correct/Incorrect Feedback

Advanced Organizer

TEST SUMMARY

Suggested Topics for Review

1. Nature of Fire
2. How to Report a Fire
3. Response to Fire Alarm
4. Compartment Identification
5. Material Conditions of Readiness
6. Breathing Devices
7. Protective Clothing
8. Performing in a Firefighting Party
9. Classes of Fire and Fire Extinguishing Agents & Systems
10. Dewatering Equipment

Number Correct: 18

Test Score: 90%

Suggested topics for review are highlighted in red.

OK

Test - Student Performance

Role Modeler

Student Tracking - Data Collection

FIREFIGHTING VALUES

Please fill in the following information:

SSAN (####-##-####)   123-45-6789

Last Name   DOE

First Name   JOHN

Gender (M or F)   m

Division

QUIT

Role Modeler

FIREFIGHTER VALUES
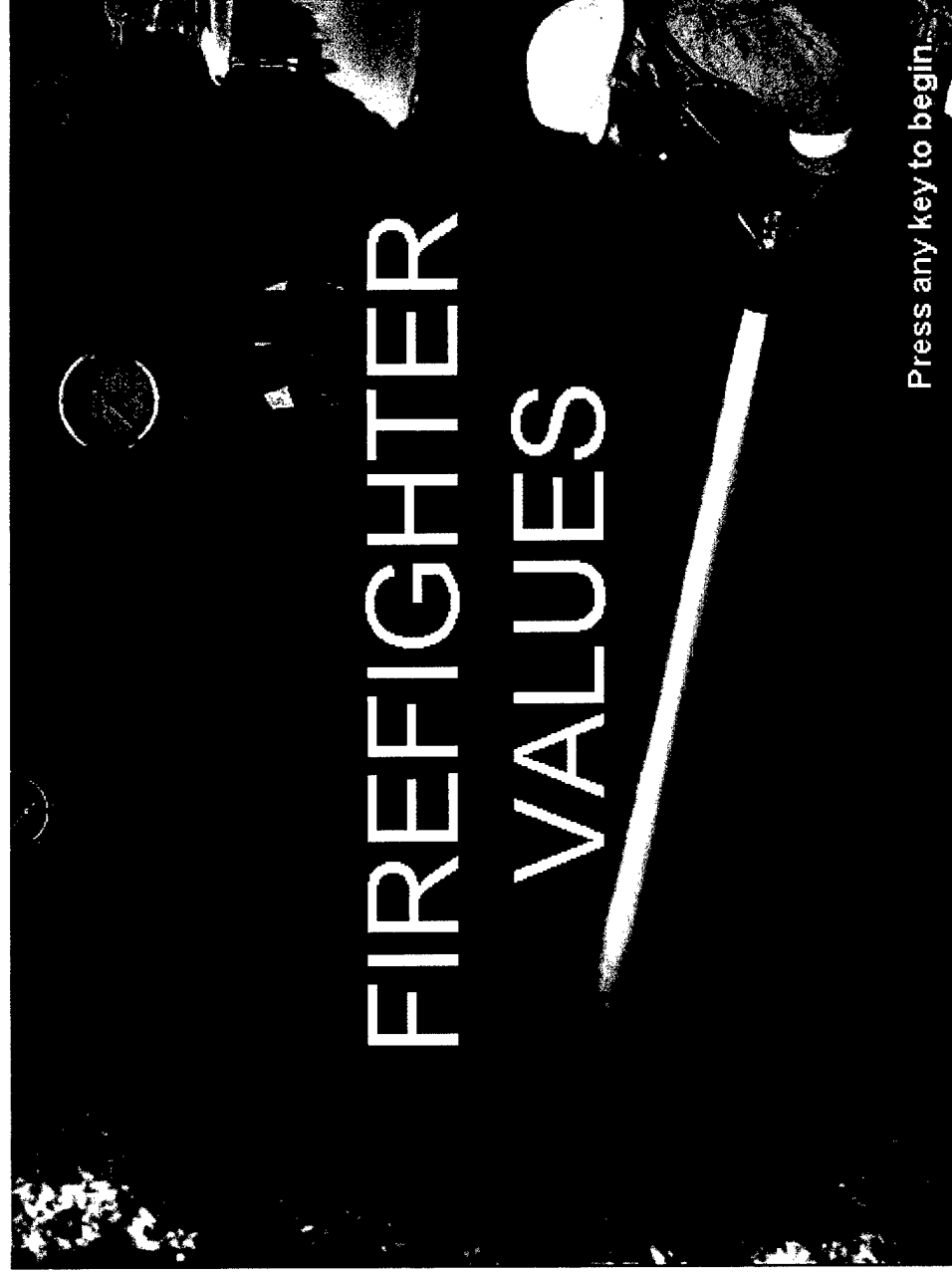
Press any key to begin...

Role Modeler

LEADERSHIP

SAFETY

T
E
A
M
W
O
R
K

Click on a button
to learn more
about that topic.

COMMUNICATION

QUIT
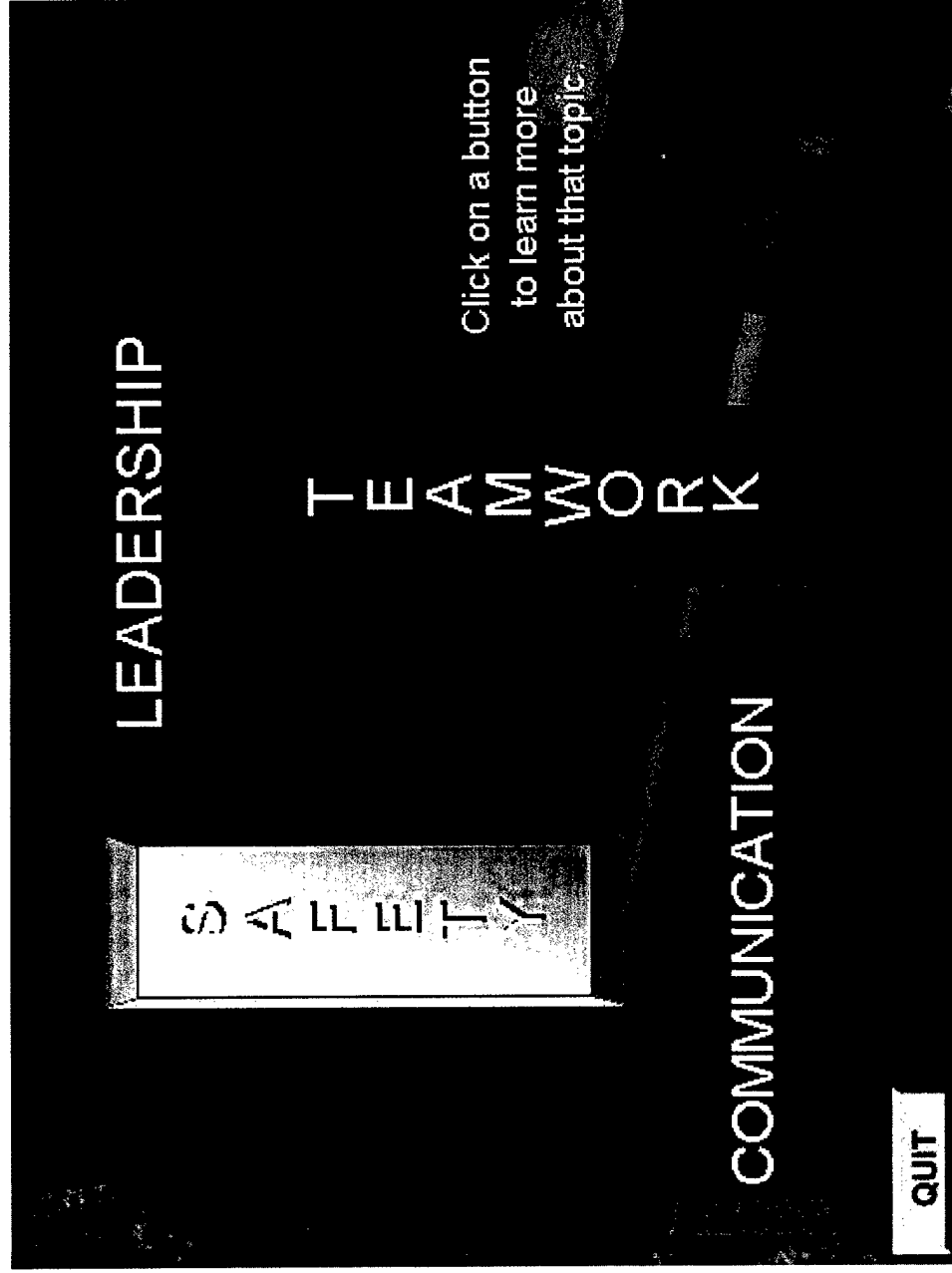
Menu

Role Modeler

Leadership Skills

L EADERSHIP

REPLAY ▶

PAUSE ||

QUIT

BACK

FORWARD
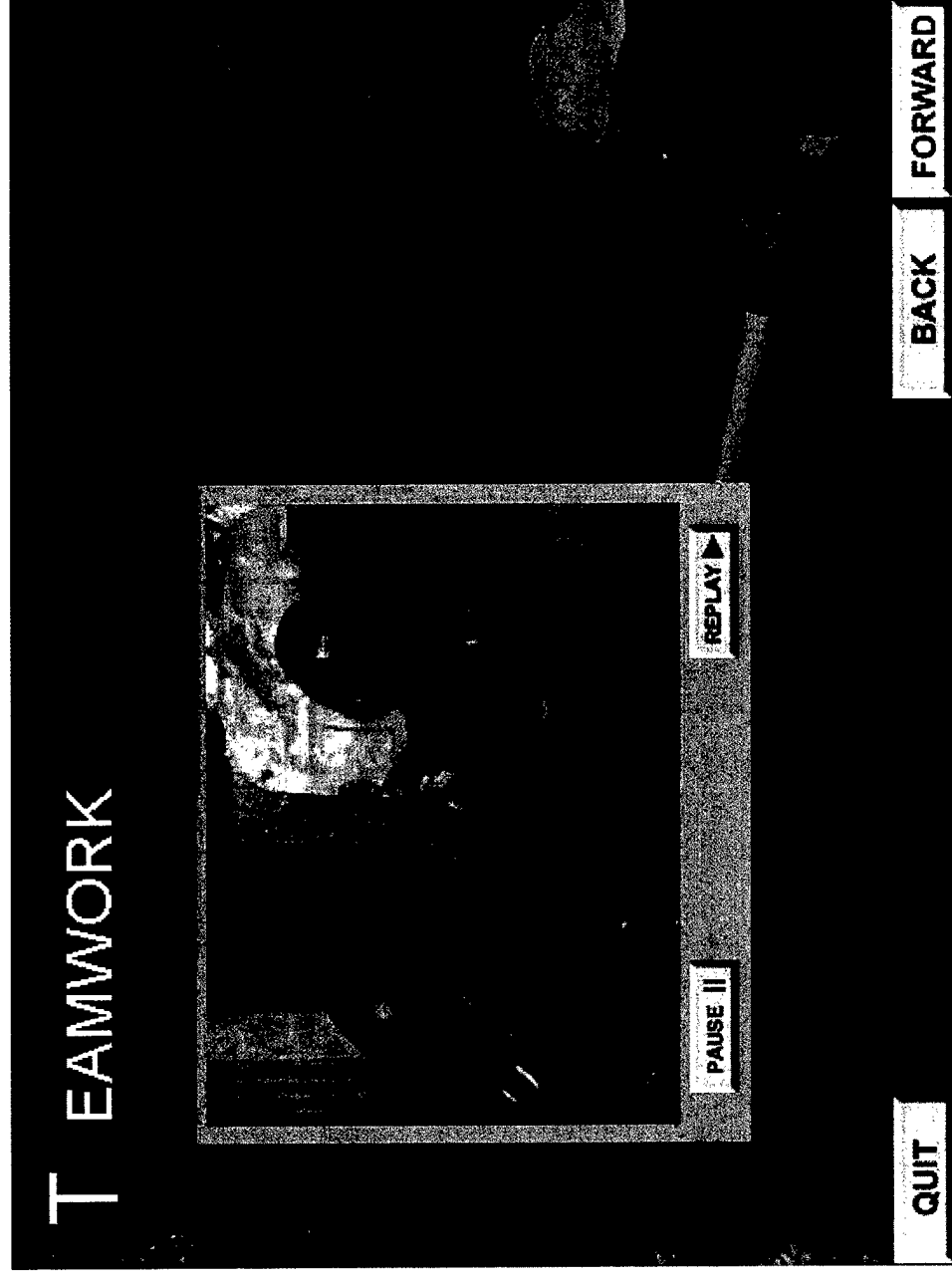
Role Modeler

Teamwork

Role Modeler

TEAMWORK

PAUSE ⏸	REPLAY ▶

QUIT

BACK | FORWARD
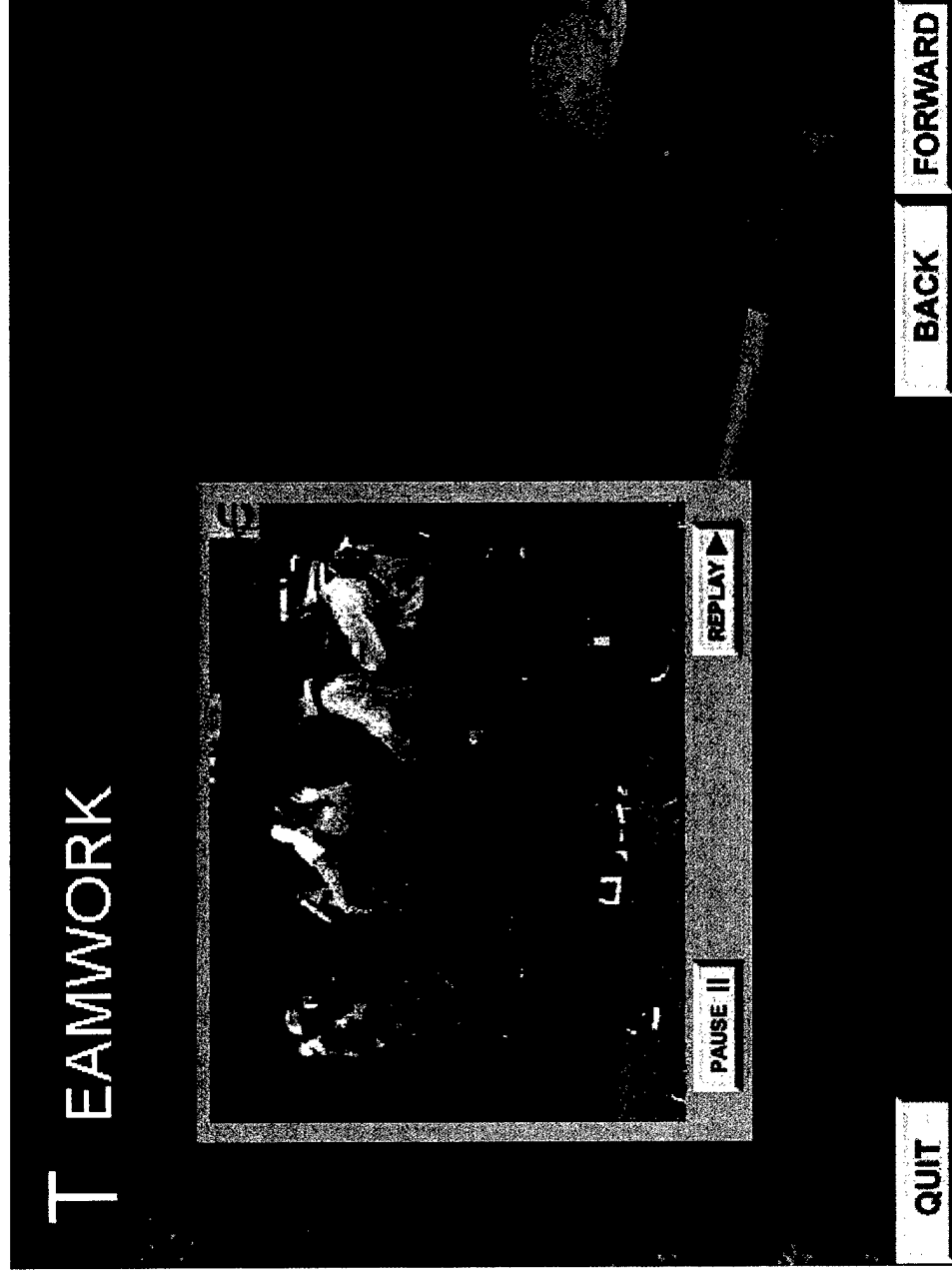
Teamwork

Role Modeler

C OMMUNICATION

PAUSE ‖   REPLAY ▶

QUIT   BACK   FORWARD
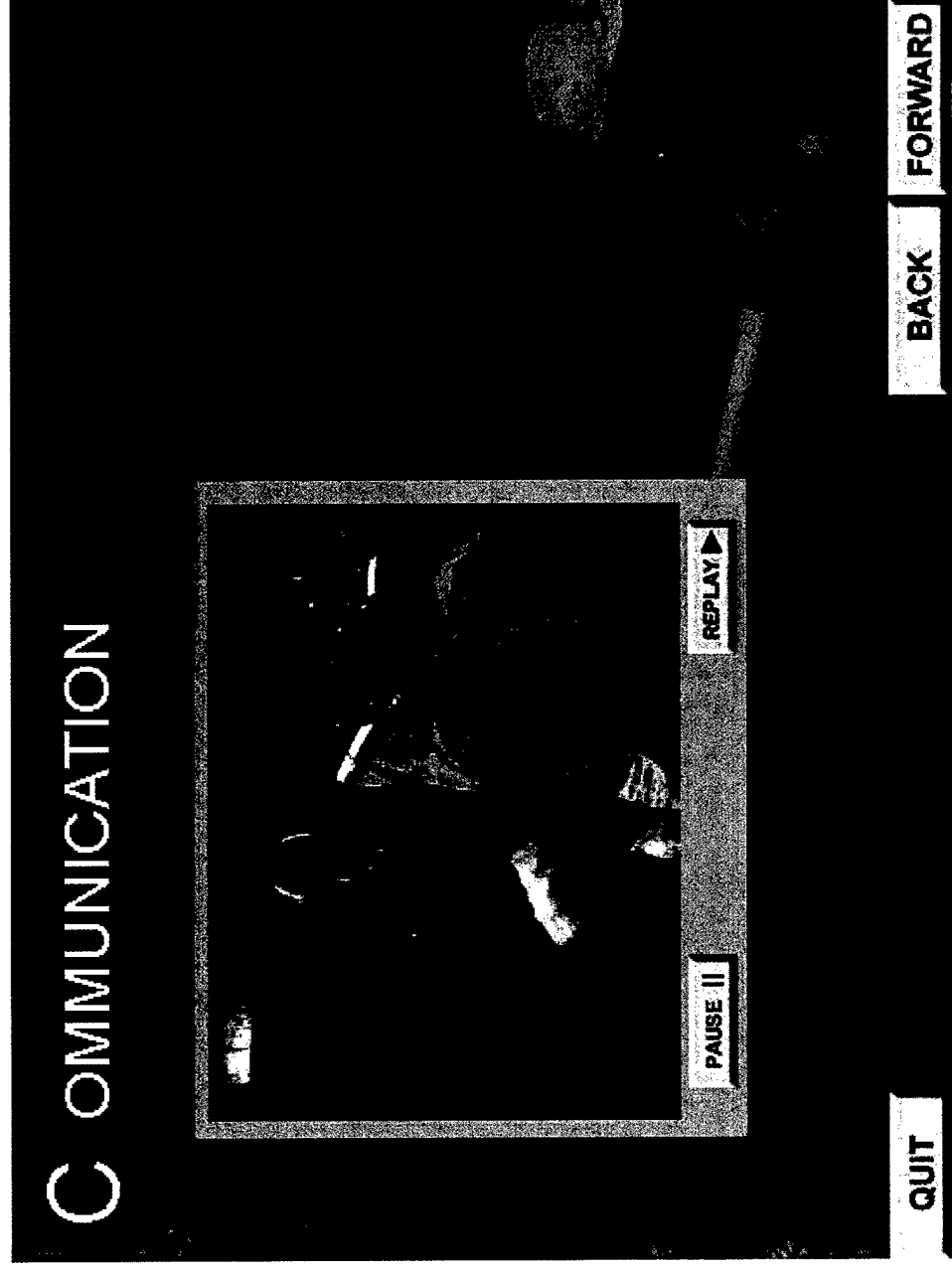
Communication Skills

Role Modeler
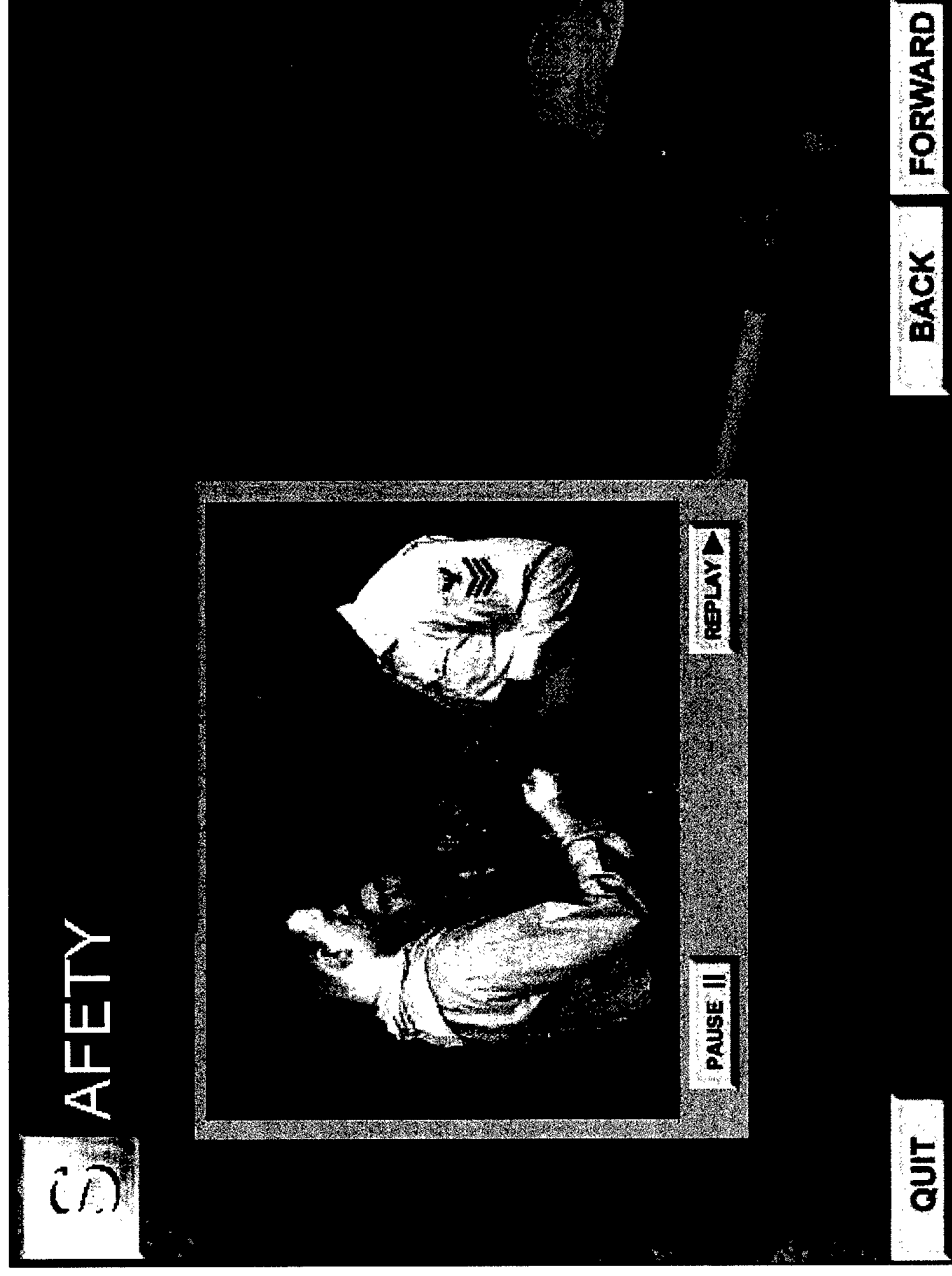
Safety Awareness

Role Modeler

TEAMWORK

SELECT THE
TEAMATES WHO
ARE SHOWING
GOOD
TEAMWORK

QUIT

BACK | FORWARD
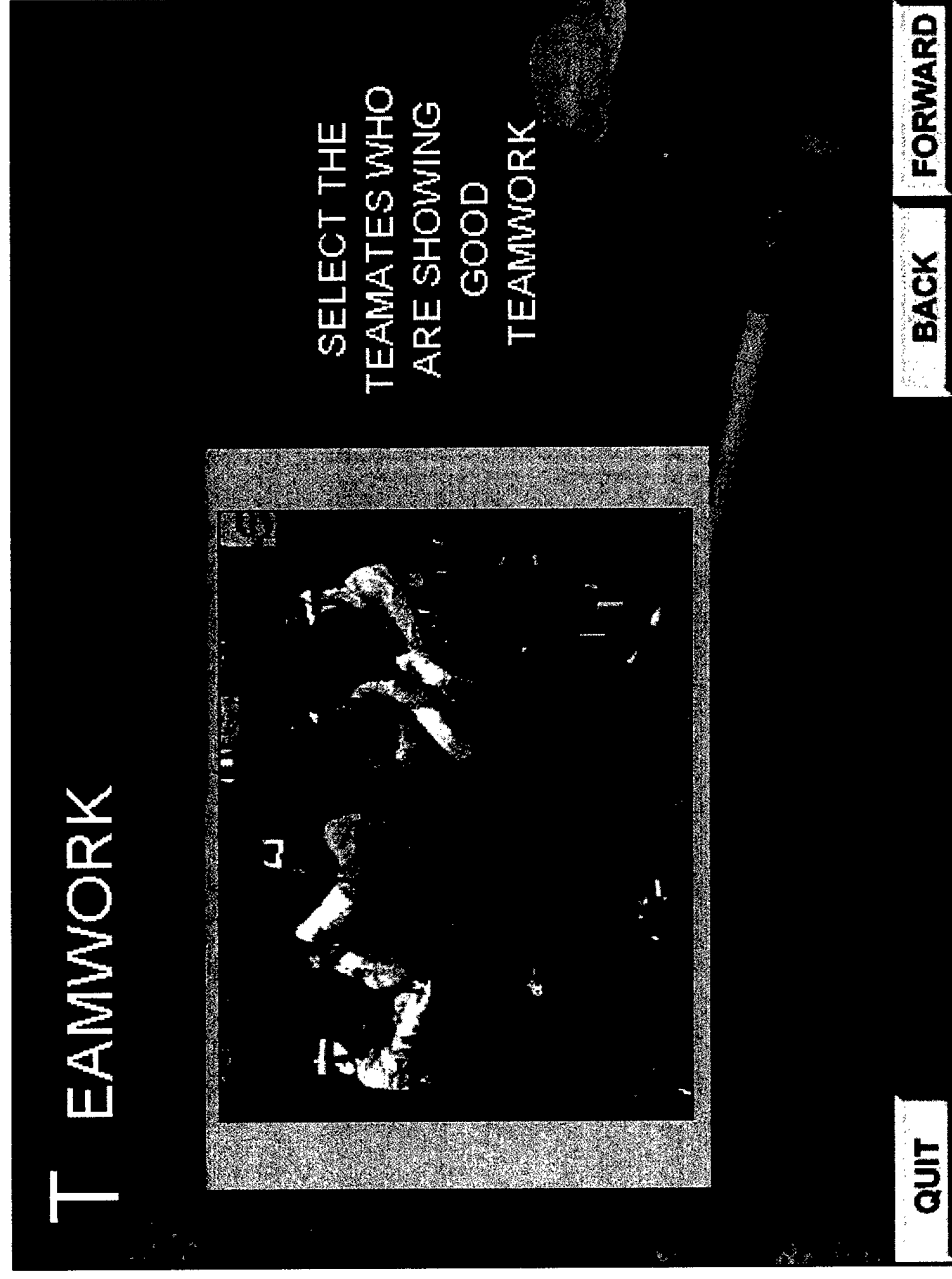
Exercises

Role Modeler

LEADERSHIP

What do you think of her explanation?

A THE TEAM LEADER IS DOING OK. THE INSTRUCTOR NEEDS TO KNOW WHO WAS MAKING THE MISTAKES.

B THE TEAM LEADER IS WRONG. THE LEADER SHOULD TAKE RESPONSIBILITY.

PAUSE || REPLAY ▶

QUIT   BACK   FORWARD

Exercises

Role Modeler

Feedback

LEADERSHIP

What do you think of her explanation?

Good Job!

...DER IS DOING OK.
...OR NEEDS TO KNOW
...KING THE MISTAKES.

...DER IS WRONG. THE
...LD TAKE
RESPONSIBILITY.

PAUSE ❚❚   REPLAY ▶

QUIT   BACK   FORWARD

Role Modeler

SAFETY

SELECT THE SAILORS TO HEAR MORE.

QUIT BACK FORWARD
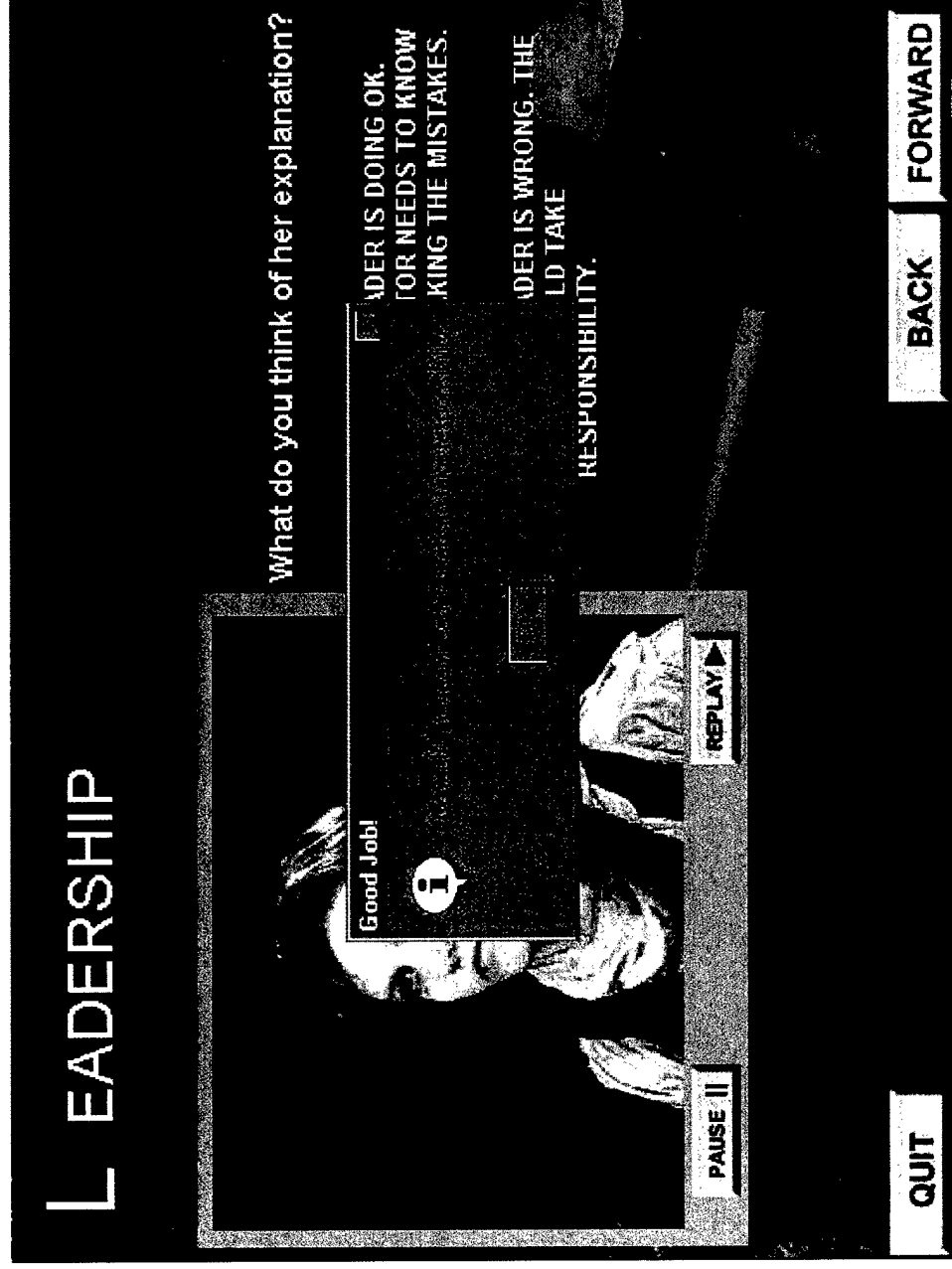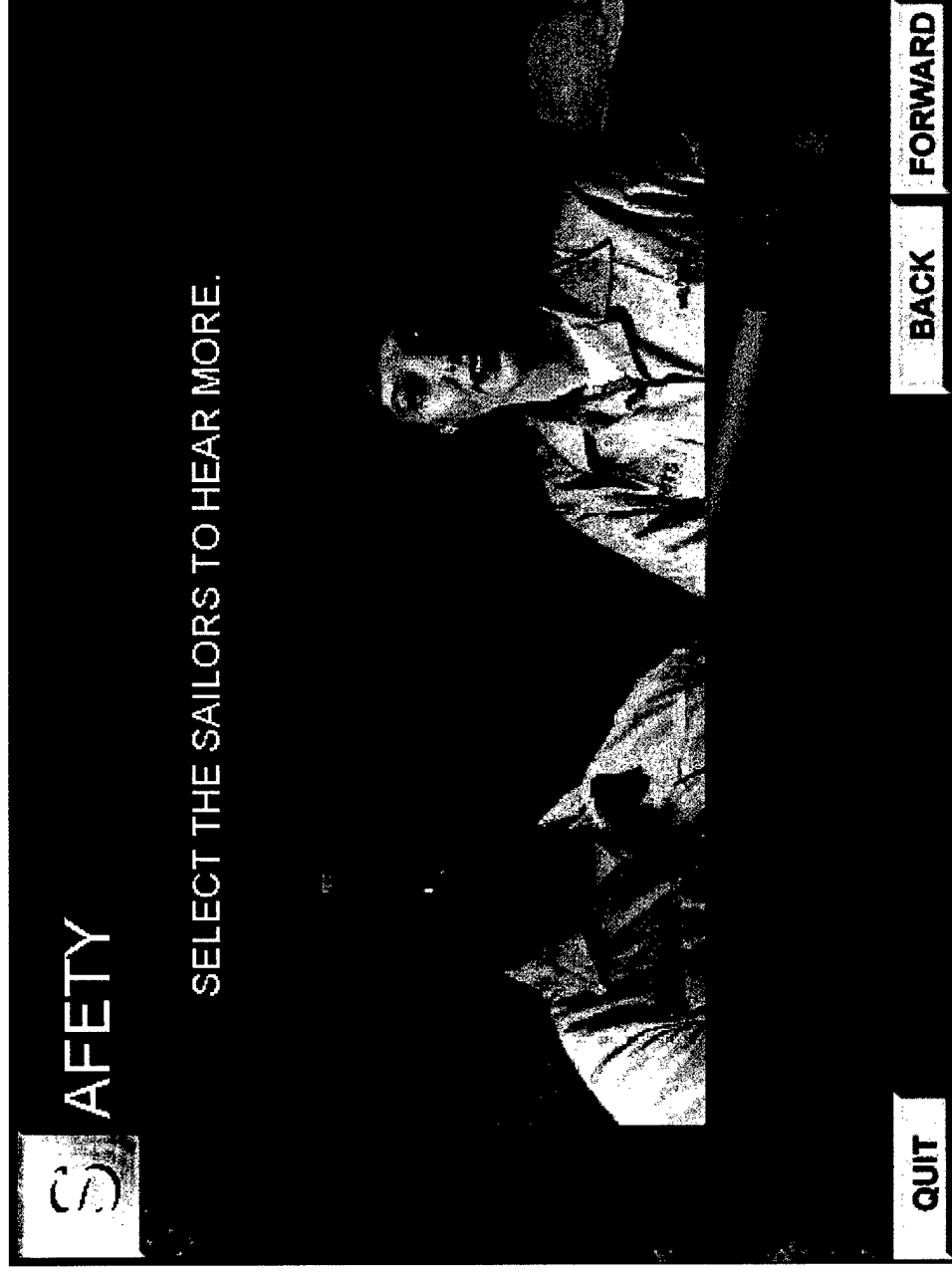
Testimonials

APPENDIX B

FINAL STRESS QUESTIONNAIRE

Recruit Questionnaire

I.D. Number - Social Security Number

Fill in your 9-digit social security number in the boxes and bubbles provided.

Test Form - Gender

Fill in your gender: A = Male   B = Female

Exam No. - Division Number

Fill in your 3 digit division number in the boxes and bubbles provided.

For each question, please fill in the bubble to describe how you see firefighting training at this time.

		[A] Strongly Disagree	[B] Disagree	[C] Neither	[D] Agree	[E] Strongly Agree
1.	I always know which tool or procedure to use.	[A]	[B]	[C]	[D]	[E]
2.	In firefighting, recruits have to do things in certain ways that could be done differently.	[A]	[B]	[C]	[D]	[E]
3.	Recruits cooperate well on firefighting tasks.	[A]	[B]	[C]	[D]	[E]
4.	Recruits always follow standard firefighting procedures.	[A]	[B]	[C]	[D]	[E]
5.	Recruits have to work very fast.	[A]	[B]	[C]	[D]	[E]
6.	I have the ability to succeed in firefighting training.	[A]	[B]	[C]	[D]	[E]
7.	I know the firefighting performance standards expected of me.	[A]	[B]	[C]	[D]	[E]
8.	All firefighting tasks are done "by the numbers."	[A]	[B]	[C]	[D]	[E]
9.	Conflicting orders from different people are common in firefighting.	[A]	[B]	[C]	[D]	[E]

		[A] Strongly Disagree	[B] Disagree	[C] Neither	[D] Agree	[E] Strongly Agree
10.	It is easy to remember the technical details of firefighting.	[A]	[B]	[C]	[D]	[E]
11.	I can overcome firefighting problems by trying harder.	[A]	[B]	[C]	[D]	[E]
12.	Team communications are clear and specific.	[A]	[B]	[C]	[D]	[E]
13.	Firefighting procedures are spelled out in detail and followed closely.	[A]	[B]	[C]	[D]	[E]
14.	You can succeed on even hard firefighting tasks if you stick with it.	[A]	[B]	[C]	[D]	[E]
15.	Firefighting rules or procedures often are in conflict, so recruits are uncertain what to do.	[A]	[B]	[C]	[D]	[E]
16.	Explanations about what has to be done are clear.	[A]	[B]	[C]	[D]	[E]
17.	Team leaders make good, quick decisions.	[A]	[B]	[C]	[D]	[E]
18.	Doing well in firefighting training is difficult no matter how hard you try.	[A]	[B]	[C]	[D]	[E]

		[A] Strongly Disagree	[B] Disagree	[C] Neither	[D] Agree	[E] Strongly Agree
19.	Any recruit who does poorly in firefighting has too little ability.	[A]	[B]	[C]	[D]	[E]
20.	Recruits often are confused over who is supposed to do which firefighting task.	[A]	[B]	[C]	[D]	[E]
21.	Success in firefighting training depends on how hard I try.	[A]	[B]	[C]	[D]	[E]
22.	It is difficult to keep up with the tasks that must be done.	[A]	[B]	[C]	[D]	[E]
23.	Team member responsibilities are clearly defined.	[A]	[B]	[C]	[D]	[E]
24.	There is only one way to do each firefighting task.	[A]	[B]	[C]	[D]	[E]
25.	My goals in firefighting training are clear.	[A]	[B]	[C]	[D]	[E]
26.	Any recruit who fails firefighting just is not trying hard enough.	[A]	[B]	[C]	[D]	[E]
27.	Success in firefighting is a matter of luck.	[A]	[B]	[C]	[D]	[E]
28.	Orders given during firefighting are clear.	[A]	[B]	[C]	[D]	[E]

For each emotion listed below, fill in the bubble that best answers the question "Do you feel __(emotion)__?"

		[A] Not At All	[B] A Little	[C] Moderately	[D] Quite A Bit	[E] Extremely
29.	Depressed	[A]	[B]	[C]	[D]	[E]
30.	Vigorous	[A]	[B]	[C]	[D]	[E]
31.	Cheerful	[A]	[B]	[C]	[D]	[E]
32.	Uneasy	[A]	[B]	[C]	[D]	[E]
33.	Blue	[A]	[B]	[C]	[D]	[E]
34.	Satisfied	[A]	[B]	[C]	[D]	[E]
35.	Angry	[A]	[B]	[C]	[D]	[E]
36.	Active	[A]	[B]	[C]	[D]	[E]
37.	Sad	[A]	[B]	[C]	[D]	[E]
38.	Annoyed	[A]	[B]	[C]	[D]	[E]
39.	Happy	[A]	[B]	[C]	[D]	[E]
40.	Energetic	[A]	[B]	[C]	[D]	[E]

APPENDIX C

LEARNING STYLES PRETEST

1. Which do you like better, instructions which are *structured for you* or learning *on your own* where *you* control the *order* of what you are learning?

Let's say you are setting up your brand new VCR. You have the instructions. Do you follow the instructions *step-by-step* or do you jump around, hooking things up in a random order until you get it put together?

A Structured/Step-by-Step

B On my own/jumping around/ random order

FORWARD

---

2. Do you like to recieve feedback even if you got the answer right, or *only* when you make a mistake?

Let's say you answered a question correctly in class. Do you like the instructor to tell you that you were right and why you were right or only give you feedback when you are wrong?

A Yes, I like feedback when I am right *and* when I am wrong.

B No, I only like feedback when I am wrong.

FORWARD

3. You have to go to a friend's house and you need instructions to get there. Would you rather look at a map or read step-by-step instructions?

A Give me a map.

B Give me instructions written out.

FORWARD

---

4. Which do you prefer, learning *details first* and then seeing how things fit together or seeing the *"big picture"* first and *then* getting all the details.

Let's say you have to learn how an automobile engine works. Do you want to know the *details* about each part of the engine *first* or do you want to know *in general* how the engine works and *then* learn the details?

A Give me the details first.

B Give me the "big picture" first.

FORWARD

5. If you had to choose *one method* for receiving instructions which would it be, pictures or words?

A Pictures.   **B** Words.

FORWARD

---

6. Which do you prefer, *being told* everthing about a topic or discovering things about the topic *on your own?*

Let's say you have to assemble a bookcase. Would you rather have a friend show you how to put it together or would you prefer to put the it together on your own.

A Follow my friend's instructions.   **B** Learn on my own.

FORWARD

7. To learn how something works, would you rather *get your hands on it* and *play with it* or would rather *think and read* about how it works?

Let's say you have to learn how a telephone works. Would you read the instructions about how it works or would you take it apart to learn how it works?

A Take it apart.

B Think and read about how it works.

FORWARD

---

8. Do you *accept* ideas as they are presented to you or do you like to *challenge* ideas and try to change them?

Let's say your history teacher is describing a historical event. Are you likely to question her?

A I accept information as it is told to me.

B I challenge information and ask questions.

FORWARD

9. Which do you like better, learning in a classroom or learning on your own?

A Classroom.   **B** On my own.

FORWARD

---

10. Which type of courses do you like better:
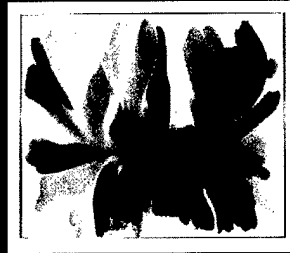
A Math and Science   **B** Art and Music

FORWARD

11. Which painting do you like best?

A Realistic   **B** Abstract

FORWARD

12. When you are learning to do a task, do other people's opinions matter to you?

A No, only my opinion matters.   **B** Yes, other people's opinions matter.

FORWARD

13. How do you learn best, by being *challenged* or by being *supported?*

Let's say you are learning how to swing a baseball bat. Would you learn best if your coach told you that you have what it takes to be a great hitter? Or would you learn best if your coach told *you* that you could *do better* than any of the other hitters.

A Challenge, can do better than the other hitters.

B Supportive, have what it takes to be a great hitter.

FORWARD

---

14. Which do you like better, projects that have a *clear structure and a set plan and goal,* or projects where *you* get to *figure out for yourself* what has to be done and how to do it?

Let's say you have to write a term paper. Would you prefer to have the instructor give you the topic and outline or would you rather come up with your own topic and outline?

A Clear structure with set plan and goal.

B Figure out for myself what has to be done and how.

FORWARD

15. Which do you prefer, doing one thing at a time or doing several things at the same time?

A   One thing at a time.   B   Several things at a time.

FORWARD

Recruit Training Command

Southwest Research Institute

APPENDIX D

RESEARCH PAPER

RESEARCH AND DEVELOPMENT OF INTELLIGENT TUTORING STRATEGIES FOR U.S. NAVAL RECRUITS

Katharine Golas, Ph.D., Claire S. Bartoli, Sherri Miller
Southwest Research Institute
San Antonio, Texas

Imelda Idar, Ph.D.
U.S. Navy

ABSTRACT

Since 1990, the attrition rate of U.S. Navy recruits has increased significantly. While psychological disorders are the most significant cause of attrition for females, males separate primarily for disciplinary and legal offenses. Naval training experts have long realized that academic difficulties are often manifested in problem behaviors that lead to separation. The Navy also recognizes that their operational and training environments are male-dominated, and there is a growing concern that gender integration in training has not been appropriately addressed. To stem attrition, maximize the acculturation process, and align training to address Fleet requirements, the Navy is developing and testing new instructional strategies and pedagogical practices.

The objective of this research effort was to increase the academic success of female recruits in a technical aspect of recruit training, where historically female recruits have performed less satisfactorily than male recruits have. A multimedia Interactive Courseware (ICW) instructional intervention called the Advanced Organizer (AO) was developed and tested in the context of firefighter training, to provide a structure for acquiring and storing technical material. The AO utilizes an intelligent tutoring strategy, where behaviors which reflect a student's cognitive learning style are identified and accommodated.

Data was collected at the Recruit Training Center in Great Lakes, Illinois, with over 1,000 recruits participating in the study. Results indicate that the AO decreased stress and made a significantly positive impact on the academic success of both male and female recruits.

ABOUT THE AUTHORS

Dr. Katharine Golas is the Program Manager of Instructional Systems at Southwest Research Institute, where she has worked since 1988. Dr. Golas has worked in the training industry for 22 years, and specializes in the development of training technologies. She has an M.S. and Ph.D. in Instructional Systems from Florida State University and a B.A. from Georgetown University.

Claire Bartoli is a Senior Research Analyst at Southwest Research Institute, where she has worked since 1995. In her 18 years of experience in the education and training industry, Ms. Bartoli has managed and participated in projects involving all phases of instructional development. She has an M.S. in Education from Wilkes University and has completed graduate work in Adult Education at Virginia Polytechnic Institute.

Sherri Miller is a Technical Specialist at Southwest Research Institute, where she has worked since 1992. Ms. Miller is an expert in Interactive Courseware (ICW) and Electronic Performance Support System (EPSS) development. She has authored numerous ICW and EPSS programs using Macromedia Authorware, Director, IconAuthor, Robohelp, and Visual Basic.

Dr. Imelda Idar has worked for the Navy since 1984. She began her career as the program officer for Recruit Training, and in 1998, assumed her present role as Deputy Director, Officer Education and Training (CNO N73C). Dr. Idar has a B.S. in Chemistry from St. Mary's University, an M.A. from the University of Texas in Bicultural/Bilingual studies, and a Ph.D. in Linguistics from Georgetown University.

RESEARCH AND DEVELOPMENT OF INTELLIGENT TUTORING STRATEGIES FOR U.S. NAVAL RECRUITS

Katharine Golas, Ph.D., Claire S. Bartoli, Sherri Miller
Southwest Research Institute
San Antonio, Texas

Imelda Idar, Ph.D.
U.S. Navy

BACKGROUND

In Fiscal Year 1997, the Navy launched extensive efforts to re-engineer training to improve its efficiency and effectiveness. The genesis of these initiatives was the significant need to man the Fleet with well-trained and disciplined warriors capable of operating the highly technical platforms. As the re-engineering training progressed, Navy recruit attrition became a major concern. Analysis of attrition factors at the Recruit Training Commands (RTC) indicated that the attrition rate for U.S. Navy recruits increased from 8.9% to 13.4%. Attrition data analysis pointed to psychological disorders as the most pronounced reason for the separation of recruits. Training experts understand that there is a high correlation between this type of attrition and learning difficulties. Accordingly, they began exploring instructional theories and paradigms that would enhance and support the re-engineering training efforts. Research efforts focused on recruit learning strategies within the Navy training context to gain greater knowledge of which training paradigms and supporting technologies would effect a greater yield in terms of optimized training effectiveness and reduced recruit attrition.

The focus of this paper is to report on the development and testing of an instructional aid entitled the Advanced Organizer (AO). The goal of the AO was to bridge the gap between what the recruits already knew and what they needed to know in the context of the firefighting environment. The AO orients recruits toward learning strategies, which parallel their cognitive learning styles. To determine the effectiveness of the AO, the research assessed basic psychometric information on stress measures, i.e., role clarity (the part a recruit plays in a firefighting team) and self-efficacy (confidence in recruits' ability to perform a task), role conflict (recruit interaction with team leaders) and negative affect (recruit mood—sadness, annoyance, depression, cheerfulness); and impact on final firefighting test scores. Analysis factored in the Armed Forces Qualification Test (AFQT), a subset of the Armed Services Vocational Aptitude Battery (ASVAB). The AFQT has a strong correlation to general intelligence. It was used as a barometer of accession quality along with the educational level.

The recruit population involved in the study reflected a cross section of the normal annual RTC population. The control and experimental groups were comprised of recruits from the fall and summer month cohorts. Demographics for the study participants are shown in Table 1.

Table 1. Distributions of Demographic Characteristics

Group	% of Sample
Age	
17 years	7.7%
18 years	53.7%
19 years	15.9%
20 years	7.9%
21-34 years	14.7%
Ethnicity	
American Indian	4.0%
Asian/Pacific Islander	5.4%
Black	25.1%
Hispanic	11.8%
White	52.7%
Education	
Less than 12 years	3.8%
12 years	91.3%
More than 12 years	4.9%

ADVANCED ORGANIZER (AO) TREATMENT

To design the AO treatment, research was conducted in the areas of cognitive learning styles (Pask, 1976; Rowland and Stuessey, 1988; Hanson, 1994; Shlechter, 1986; Kolb, 1976; Messick, 1984; Tannen, 1994; Jonassen and Grabowski, 1993) and learner control (Daniels, 1996; Frey and Simonson, 1994; Merrill, 1984; Williams, 1993; Reigluth and Stein, 1983). The research findings were used as a basis for deriving the instructional strategies on the following premise: to overcome difficulties with the content of technical material, represented by the academic component of firefighter training, not only do recruits need analogies they can understand, they need an introduction to the information in a way that complements their unique learning styles.

The goal of the AO treatment was to bridge the gap between what the recruits already knew and what they need to know in order to meaningfully learn to fight fires. The AO treatment provided a context of meaning for new information to be learned. It was designed to help orient the students to the firefighting subject matter in such a way that the subject matter was directly related to any pre-existing knowledge students may have. It was expected that the AO would help the students anticipate the performance requirements of the job by letting them know what to expect, as well as demonstrate the desired behaviors and attitudes for acceptable job performance. The overall objective of the AO was to provide recruits with a context of knowledge that would prepare them to succeed in the academic and hands-on components of firefighter training.

One of the goals for the design of the AO was to present a conceptual framework for the new knowledge and skills presented in firefighter training, a framework that would appeal to different learning styles. In designing the strategies, two broad categories of learners were created from the different cognitive strategies. Various instructional strategies, expected to be effective for different learning styles, were applied to each of the two broad categories of learners. They were applied as prescriptions for strategies that would present the material most effectively for each type of learner. The instructional strategies supporting the two learning styles are described in Table 2. This does not imply that a learner is expected to use one cognitive strategy in all situations. We strongly suspected that learners would switch strategies to accommodate different subject areas and different circumstances. Therefore, the AO provided different instructional modes to accommodate different cognitive strategies, and recorded which strategies were being used at particular points in the instruction.

Table 2. Instructional Strategies for the AO

Field-Dependent Learners	Field-Independent Learners
Provide structured instructional sequencing.	Provide learner control features.
Provide continuous performance feedback.	Provide necessary feedback.
Provide a verbal overview.	Provide a graphic overview with complex graphics.
Structure information to be presented.	Allow students to derive the structure of the information.
Associate ideas with their practical application.	Show (graphically) why the material is important.
Provide explicit information about performance outcomes.	Allow students to inquire about performance outcomes.
Explain how the various components fit together, then provide a global perspective.	Show the performance (global perspective) and let students identify the components.
Provide a verbal organizer.	Provide a graphic organizer.
Explain why damage control is important and how each individual contributes to the welfare of all.	Explain why it is necessary to acquire all the knowledge and skills to perform effectively.
Provide a verbal description of the "big picture."	Use a graphic presentation to show the "big picture" and let students derive the specifics.
Provide "context" for the roles the learners will perform as part of the team and what they will need to know to fulfill their responsibilities.	Provide "context" in terms of what they will learn and how they will use the knowledge to perform their roles in the firefighting team.
Use an expository presentation: Present a generality and demonstrate the necessary skills to understand the higher-level skill.	Use a discovery presentation; allow the learner to discover the higher-level skill.
Emphasize that during training, "hands-on" experiences will be provided.	Emphasize that during training, learners will use knowledge gained in academic training.
Provide information from the specific to the general.	Provide information from the general to the specific.
During demonstrations, emphasize "feeling" aspects of activities: working toward the common good.	During demonstrations, emphasize "thinking" aspects of activities; develop expertise.

3

The AO used a combination of learner control and intelligent tutoring, with assessment of student requirements and presentation of appropriate material, to provide a unique instructional experience to each student. The system presented information in ways that specifically addressed the style and/or concerns of the learner.

Learner Control

In addition to the typical learner control features such as pacing and review, the learner can select two different expository approaches: a structured instructional approach (linear), or an approach in which the learner imposes his or her own structure (concept map). Given the learning preferences indicated in the research, it was expected that a field-dependent learner would prefer the linear approach, while a field-independent learner would prefer the concept map approach.

Selection of Linear Approach

If the student selected the linear approach, this indicated his or her preference for structured instructional sequencing and a verbal overview of the topic. Since the student had exhibited characteristics of a field-dependent learner, the system structured the lesson for the student from specific to general.

Selection of Concept Map

If the student selected the concept map, this indicated his or her preference to have control over the sequencing of information, thus exhibiting characteristics of a field-independent learner. Since the student had indicated that he or she was a field-independent learner, the system provided a graphic overview of firefighting in the form of a multimedia concept map. Once the introduction was over, the student was able to select specific topics in any order. In doing so, the student was deriving his or her own structure of the information and exercising learner control over the structure and sequencing of the information. The concept map structure fits the field-independent learning style by providing a "big picture" of the lesson material before the student's attention turns to the details of the subject.

Learner Questions

The student can ask the system questions which are provided on each expository screen in the AO, regardless of whether the user is accessing the screen via the linear approach or the concept map approach. The purpose of the questions is twofold. First, they provide individualized, personalized instruction within the context of the computer-based training environment. Questions and answers are specific to the content in the lesson. Second, they provide a basis for automatically tailoring the information presented in the lesson to the user, based on the computer's assessment of the user's concerns. There are four categories of questions: gender, information, role, and remember. If the user asks more than one question in a category as he or she proceeds through the presentation, additional information relevant to that category is automatically presented for subsequent content areas. For example, the user may indicate, by asking questions in the gender category, that he or she is concerned about whether females can be effective firefighters. After the user has asked two "gender" questions, information on females in firefighting is provided in subsequent content areas.

Table 3 shows the four categories of questions, the sample questions, the instructional concerns which the questions are designed to address, and the purpose of providing answers and additional information in the AO treatment.

The learner questions are answered using multimedia presentations with motion video, audio, still photographs, and text, then the presentation is continued. Throughout the instruction the learner retains control of the pace and presentation. The learner can take the test at any time. The average recruit took 55 minutes to complete the AO treatment.

4

Table 3. AO Learner Questions

Category	Sample Question	User Concern	Purpose
Gender	Is the weight of the equipment a problem for female sailors?	Recruit is showing concern about the ability of females to perform successfully in firefighting.	Information is designed to bolster confidence in males and females that females can be successful at firefighting.
Information	How is this like something I already know? Can you explain the ID system further?	Recruit is asking for additional information specific to the topic. In some cases the recruit is asking the program to help relate new information to known information.	Information is designed to help users who need additional tutoring to understand concepts.
Role	Why is this information important to me?	Recruit is exhibiting a concern for how the activities relate to the common good. This is a typical concern for a field-dependent learner.	Information is designed to provide users with information that will help them understand their role in firefighting teams.
Remember	What exactly do I have to remember about this topic?	Recruit wants to know exactly what information is important to remember. This is a typical concern for a field-independent learner.	Information is a synopsis of the content area, which the user can use as a mental checklist for what is important to remember.

Learning Style Pretest

A learning style pretest was designed to provide specific data regarding which instructional strategies were the most successful at preparing students for the academic portion of the course. The learning style pretest was computer-based and contained 15 items which determined the broad category of learning style for the student (see categories of learners described in Table 2).

PROCEDURES FOR DATA COLLECTION

Development of CD-ROMs for Each Instructional Strategy

The AO program was organized into six separate programs, one for each of the six instructional strategies. The six instructional strategies or "routes" through the AO were:

- Concept Map/No Questions. In this program, the students were in the concept map approach and they were not able to ask the computer any questions.

- Concept Map/All Questions. In this program, the students were in the concept map approach and they were provided with all the questions and answers that were possible in the program.

- Concept Map/Student Selected. In this program, the students were in the concept map approach and they were able to ask questions if they wanted to.

- Linear/No Questions. In this program, the students were in the linear approach and they were not able to ask the computer any questions.

- Linear/All Questions. In this program, the students were in the linear approach and they were provided with all of the questions and answers available in the program.

- Linear/Student Selected. In this program, the students were in the linear approach and they were able to ask the computer any questions.

Preparation of Classroom-Based AO

The control group in the data collection effort was provided with one hour of classroom instruction covering the basic firefighting content of the multimedia AO. The instructor at Great Lakes prepared this one-hour block of classroom instruction prior to the data collection effort.

Academic Firefighting Test

During the data collection effort, the control and treatment group students were given the academic firefighting test twice. The first time they took the test was immediately upon completion of the multimedia AO or one-hour classroom instruction. The second time they took the test was upon completion of the firefighter training course, which lasted five days.

Firefighting Stress Questionnaire

A 40-item stress evaluation questionnaire was developed and administered to the students prior to and after firefighter training. The questionnaire measured self-efficacy and perceived stress. The Committee for Protection of Human Subjects approved the protocol to administer the stress questionnaire and instructional treatment to the recruits.

DATA ANALYSIS

Data was collected over a three-year period at the Recruit Training Center in Great Lakes, Illinois. One thousand two hundred and thirty-five (1,235) recruits participated in the study.

The analysis approach consisted of five steps. First, the relationships between potential determinants of academic performance were examined. The objective was to determine whether there was significant risk of confounding the effects of one determinant (e.g., gender) with that of another determinant (e.g., field dependence).

The second step in the analysis created several residualized performance variables. Raw scores on performance tests were adjusted to take into account differences in intelligence as measured by AFQT.

The third step in the analysis examined gender differences in anticipated stress, and assessed the differences between anticipated stress measured at the beginning of firefighter training and the actual

stress perceived at the end of firefighting. Independent and paired sample t-tests were performed.

The fourth step in the analysis examined the effects of AO design. This part of the analysis compared the different AO types (linear/concept mapping) and the use of questions (none/student-selected/all) in what was basically a 2 x 3 analysis of variance (ANOVA). The key issue in this part of the analysis was the determination of whether particular elements of AO design affected performance. The basic ANOVA in this step was extended to include gender (male/female) and learning style (field-dependent/independent) as group classification variables. These extensions made it possible to determine whether the effect of AO design depended on the recruits' gender and learning style. The final ANOVA design, therefore, was a 2 x 2 x 2 x 3 group classification based on gender, learning style, AO type, and question utilization.

The fifth step in the analysis compared the AO to the control group. Gender and learning style were retained as variables to determine whether the general effect of the AO, in contrast to specific effects of particular types of AO, depended on these variables.

RESULTS

This study examined four variables that might explain differences in firefighting performance: gender, AFQT, learning style, and experimental treatment. The study also examined anticipated stress. The results of the study are as follows:

a. Males and females were evenly distributed across treatment groups ($\chi^2 = 2.06$, 7 χ, p = .956).

b. AFQT scores were comparable across the treatment groups ($F_{7,271} = 0.49$, p = .841).

c. Learning style was mildly confounded with treatment group ($\chi^2 = 10.05$, 5 df, $p = .074$), largely because a higher rate than expected of field-dependent individuals were in the linear/all questions AO condition (65.5% vs. 40.9% in the other conditions).

d. The average male had a slightly higher AFQT score (62.1) than the average female (58.5), but the difference was statistically trivial (point biserial $r = -.09$, $t_{277} = 1.58$, $p = .116$).

e. Females were more likely to be field-dependent than males (60.5% vs. 51.4%), but the trend was not significant ($\phi = .09$, $\chi^2 = 1.50$, 1 df, p = .221).

f. On the first academic firefighting test, gender produced a significant main effect on adjusted performance ($F_{1,148} = 4.11$, $p = .045$). Men scored higher than women (Male average = .23; Female average = -.19).

g. On the first academic firefighting test, two interactions were statistically significant: AO type x gender ($F_{2,148} = 5.58$, p = .005) and question utilization x learning style ($F_{2,148} = 3.41$, p = .036).

h. On the second firefighting post test, where academic test performance was adjusted for AFQT, the main effects of AO type and question utilization were nonsignificant.

i. On the second firefighting academic test, where academic test performance was adjusted for AFQT and the first firefighting test performance, AO type and question utilization main effects were weak ($\epsilon2 < .014$) and statistically nonsignificant (p > .407).

j. The AO decreased stress. Recruits who completed the AO reported greater role clarity and self-efficacy at the end of firefighter training.

DISCUSSION

Effect of AO on Stress

Two results merit special attention with regard to stress in the context of the application of firefighting. The first point is regarding communication. Stress arises in part when communication lacks clarity and consistency (Kahn, et al., 1964; King and King, 1990). Particularly in dynamic teamwork situations, such as those likely to occur in firefighting, the process of communication and coordination plays a key role in determining the stress level. Thus, communication problems such as those outlined by Tannen (1994), and gender-driven behavior, as detailed by Gray (1987) become more pronounced. One objective of the AO, which was achieved, was improved communications. The experimental group had improved role clarity and self-efficacy.

The second point is that role ambiguity and role conflict tend to be positively correlated when people are asked to describe their jobs. If this trend is generalized to firefighting as a specific element of sailors' jobs, measures of these stresses may be correlated. In the present research, the hypothesis was posited that female recruits would exhibit more stress than males since females were being trained in a environment that has been historically all male. Again, role clarity and self-efficacy were improved for the experimental group. However, the assumption that females were more stressed than males was not supported. Male and female recruits were shown to be equally stressed.

Effect of AO on Recruit Performance

With regard to recruit performance in firefighting, the following points summarize the results.

A course overview helps with academic performance. Both the multimedia AO and the classroom/instructor AO improved performance significantly over standard instruction. This finding supports the premise that in order to overcome difficulties with the content of technical material, represented by the academic component of firefighting, not only do students need analogies they can understand, they also need an introduction to the information in a way that will help them bridge the gap between what they know coming into firefighter training and what they need to know to pass the course.

The format of the AO was more important in the short term than over the full duration of the course. The multimedia AO produced significantly better performance on the test immediately following the treatment than the classroom/instructor AO. However, the recruits given the classroom/instructor AO "caught up" by the end of the course and did as well on the final firefighting test as the multimedia AO group. However, even though the classroom/instructor and multimedia AO's had the same end result on performance, the multimedia AO may be more efficient. The multimedia AO significantly improved immediate performance relative to the control, which indicates there was less learning for the AO recruits during the subsequent formal instruction.

This inference follows from the equivalent level of final performance. If final performance reflects an asymptote determined by the nature of the material and the abilities of the student, the initial advantage to the multimedia AO suggests it approached this asymptote more rapidly than the classroom/instructor AO. Both methods appear to raise the asymptote, based on comparison to the standard instruction controls. Perhaps a series of multimedia instructional tools would significantly shorten the total required instructional time.

7

This question cannot be resolved in the present design for two reasons. First, more interim measures of learning would need to be developed to fully define the learning curve. Second, additional multimedia instructional tools would need to be developed to determine whether this approach was superior to standard instruction after the initial introduction.

Instructional strategies had some impact on the effectiveness of the AO. The combination of concept map and student-selected questions provided the most opportunity for the student to just "wander around" through the information. Recruits who were in the concept map/student-selected AO strategy did not perform as well on the academic firefighting test. This indicates that it may be important to avoid giving the recruits too much freedom. This finding should not be given too much weight until it is replicated.

Learning style may be important if trainees are forced to view all questions and answers. The overall pattern of differences was clear. Forcing the student to listen to all the questions and their answers improved performance for field-dependent recruits, but tended to impair performance for field-independent recruits. This finding supports previous studies which showed that field-independent learners do not adhere as well to externally imposed structures (McGee, 1979). They are more capable of developing their own internal referents and do not need external referents, such as the AO questions, to process information. On the other hand, field-dependent learners were shown to do better with an externally imposed framework, and external referents (McGee, 1979).

Effect of Gender. The study supported previous research findings that showed that females are more likely to be field-dependent learners, and males are more likely to be field-independent learners (Witkin, 1981). However, exploration of gender issues was not critical to the effectiveness of the AO, and learning style effects were the same for females and males.

CONCLUSIONS

Because the AO decreased stress and improved recruit performance, the application of the instructional aid to firefighter training was supported. The implication of these results is significant for future Navy curriculum development and for the professional development of Navy instructors. Communication and role definition should become major items in instructor professional development. Development of instructional aids such as the AO should be continued as an instructional approach for training technical information such as firefighting.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). NY: Academic Press.

Daniels, H.L. (1996). Interaction of cognitive style and learner control of presentation mode on a hypermedia environment. *Doctoral dissertation.* Blacksburg, VA: Polytechnic Institute and State University.

Frey, D., & Simonson, M. (1994). Assessment of cognitive style to examine students' use of hypermedia within historic costume. *Home Economics Research Journal*, 21(4), 403-421.

Gray (1987). The effect of sequence control on computer-assisted instruction. *Journal of Computer-Based Instruction*, 14(2), 54-56.

Hanson, J.W. (1994). Student cognitive styles in postsecondary technology programs. *Journal of Technology Education*, 6(2), 1045-1064.

Hays, W.L. (1963). *Statistics for Psychologists*. NY: Holt, Rinehart, Winston.

Jonassen, D.H., & Grabowski, B.L. (1993). Handbook of individual difference. *Learning and Instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kahn, R.L., Wolfe, D.M., Quinn, R.P., Snoek, J.D., & Rosenthal, R.A. (1964). *Occupational Stress: Studies in role conflict and ambiguity*. New York: Wiley.

King, L.A., & King, D.W. (1990). Role conflict and role ambiguity: A critical assessment of construct validity. *Psychological Bulletin*, 107, 48-64.

Kolb, D.A. (1976). *Learning style inventory technical manual*. Boston, MA: McBer and Co.

McGee, M.G. (1979). *Human spatial abilities: Source of sex differences*. New York: Praeger.

Merrill, M.D. (1984). What is learner control? In C.R. Dills (Ed.), *Instructional development: The state of the art, II, (pp. 221-242)*. ERIC Document Reproduction Service No. ED 298 905.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist*, 19(2), 59-74.

Pask, G. (1976). Styles and strategies of learning. *Journal of Educational Psychology*, 46, 128-148.

Reigluth, C.M., & Stein, F.S. (1983). The elaboration theory of instruction. In C.M. Reigluth (Ed.), *Instructional design theories and models* (pp.335-382). Hillsdale, NJ: Lawrence Erlbaum.

Rowland, P. & Stuessey, C.L. (1988). Matching mode of CAI to cognitive style: An exploratory study. *Journal of Computers in Mathematics and Science Teaching*, (7)4, 36-40.

Shlechter, T. (1986). *An examination of the research evidence for computer-based instruction in military training*. Technical Report 722.

Tannen, D. (1994). *Talking from 9 to 5*. New York: Avon Books.

Williams, M.D. (1993). *A comprehensive review of learner control: The role of learner characteristics*. Paper presented at the annual convention of the Association for Educational Communications and Technology, New Orleans, LA.

Witkin, H.A., & Goodenough, D.R. (1981). *Cognitive Styles: Essence and Origins*. New York: International Universities.